Statistical physics approaches to subnetwork dynamics in biochemical systems

B Bravi[1]‡ **and P Sollich**[1]

[1] Department of Mathematics, King's College London, Strand, London, WC2R 2LS UK

E-mail: barbara.bravi@epfl.ch, peter.sollich@kcl.ac.uk

Abstract. We apply a Gaussian variational approximation to model reduction in large biochemical networks of unary and binary reactions. We focus on a small subset of variables (subnetwork) of interest, e.g. because they are accessible experimentally, embedded in a larger network (bulk). The key goal is to write dynamical equations reduced to the subnetwork but still retaining the effects of the bulk. As a result, the subnetwork-reduced dynamics contains a memory term and an extrinsic noise term with non-trivial temporal correlations. We first derive expressions for this memory and noise in the linearized (Gaussian) dynamics and then use a perturbative power expansion to obtain first order nonlinear corrections. For the case of vanishing intrinsic noise, our description is explicitly shown to be equivalent to projection methods up to quadratic terms, but it is applicable also in the presence of stochastic fluctuations in the original dynamics. An example from the Epidermal Growth Factor Receptor (EGFR) signalling pathway is provided to probe the increased prediction accuracy and computational efficiency of our method.

Keywords: Model Reduction, Biochemical Networks, Intrinsic and Extrinsic Noise, Variational approximation

‡ Current affiliation: Institute of Theoretical Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

1. Introduction

Protein-protein interaction networks, such as the ones building up signalling pathways, can contain thousands of reacting species and the underlying mechanisms are still largely unclear [1, 2], with mathematical descriptions involving large systems of coupled nonlinear differential equations or their stochastic analogues such as Chemical Langevin Equations [3]. This motivates the search for approximate descriptions and model reduction techniques (see e.g. [4–7]). The aim is to simplify the overall analysis, to facilitate the interpretation of experimental data and ultimately to extract the maximal information from them. Analytical rather than numerical approximations can be particularly useful here by delivering qualitative insights.

The challenge, then, is to find efficient approximations of probability distributions over the temporal trajectories (paths) of a system with stochastic dynamics. Variational methods achieve this by optimizing an approximating distribution within some chosen family of distributions. Eyink [8] proposed a general scheme to produce approximations by applying a variational principle on a non-equilibrium effective action; see also [9, 10] for applications of this variational approach based on Poisson distributions.

The problem of approximating intractable distributions is also a key one in the area of machine learning techniques [11]. Variational approximations are there often based on minimizing a Kullback-Leibler divergence [12] to obtain the most faithful approximation. Families of approximating distributions include the mean field-type (factorized) ansatz [13, 14] or Gaussian distributions [15, 16], the latter allowing one to retain correlation information beyond mean field.

The Gaussian assumption can be viewed as a second order moment closure scheme, as first outlined by Whittle [17]. Several precedents exist for its use in modelling biochemical reactions [18, 19]; for performance comparisons to other approximations see also [20–22].

The setting we consider in this paper is the following: we consider a network of unary and binary biochemical reactions and focus on a subset of molecular species, the "subnetwork". We view this as embedded in the remainder of the larger network, the "bulk". The subnetwork is taken as given from the outset. It may be determined by which molecular species can be monitored experimentally, or which are better characterized theoretically, or indeed which are more relevant for the biological features of interest. Our aim is to calculate the time evolution of subnetwork variables without tracking the evolution of the bulk network, starting from just some prior information, in the form of the statistics of the bulk initial conditions, and the subnetwork-bulk and bulk-bulk interactions. We use a Gaussian Variational Approximation (GVA) of the path distribution for the stochastic equations describing the entire network, and then add nonlinear corrections.

We focus on the "forward" problem of modelling the dynamics of a subnetwork that is part of a known larger network. However, this also gives qualitative insights into the situation where only the subnetwork is known and the bulk of the network is thought of as providing the significant environmental fluctuations often detected in *in vivo* conditions. Deriving models for subsystem dynamics in a *systematic* and *principled* way can then ultimately also help with the "inference" problem of estimating properties of the environment from observed subnetwork dynamics.

Projection methods are one established approach to model reduction [23–26], having been originally introduced in irreversible statistical mechanics [27]. Projection approaches generically yield a reduced dynamical description for the subnetwork that includes a memory term and extrinsic noise known as "random force" (Sec. 2.1), though finding explicit expressions for these is challenging. Rubin et al. [26,28] achieved this for the setting of unary and binary reactions with mass action kinetics we consider here (with a later extension to Michaelis-Menten kinetics [29]) in the limit where the dynamics of the entire network is deterministic. These results will provide a useful point of comparison for our approach.

Our method, like that of [26], is applicable to *arbitrarily chosen* subnetworks, while other coarse-graining procedures specifically tailor the choice of subnetwork, e.g. to merge or eliminate fast reactions [30]. However, instead of employing projection operators, we derive the reduced subnetwork dynamics by integrating out or *marginalizing* the bulk degrees of freedom (d.o.f.) from a path integral representation of the full dynamics. This marginalization approach has been used also in recent works studying fixed or fluctuating random environmental conditions [31, 32]. In our treatment, the environment (bulk) is itself a network, with its own structure and dynamics. This makes the approach more flexible, allowing one for example to consider

different splits of a given network into subnetwork and bulk as might be relevant for different datasets.

We start in Sec. 2 by introducing the model for the network dynamics and the structure of its subnetwork-reduced version expected from projection methods. In Sec. 3 we outline a variational derivation of a Gaussian approximation over paths [15]. We then marginalize to obtain the approximate subnetwork dynamics, including the memory and extrinsic noise due to the interaction with the bulk (Sec. 4). To go beyond the effectively *linearized* dynamics implied by the Gaussian approximation, we write the subnetwork dynamics in terms of an "effective" (i.e. with the bulk integrated out) action and perform a perturbative expansion from which the *nonlinear* memory and noise can be read off (Sec. 5.1). The numerical solution of the resulting subnetwork equations is discussed in Sec. 5.2 and is more computationally efficient for large bulk networks than for the projected subnetwork equations of [26]. Nevertheless, as we show in Sec. 5.3, the two methods are equivalent in the deterministic limit, to the quadratic order that both treat systematically.

In Sec. 6.1 we use a toy model that is instructive in its simplicity, in order to illustrate our method and compare its accuracy to existing approximation schemes. Finally, in Sec. 6.2, we apply our model reduction to the protein-protein interaction network around Epidermal Growth Factor Receptor (EGFR), where it yields increased prediction accuracy compared to projection methods [26].

2. Set-up

We consider a reaction network of N molecular species such as proteins and protein complexes, with concentrations $\boldsymbol{x} = (x_1, \ldots, x_N)$. These are assumed to evolve in time according to a set of Chemical Langevin Equations (CLE) [3]

$$\frac{\partial x_i(t)}{\partial t} = \Phi_i(\boldsymbol{x}(t)) + \xi_i(t) \qquad i = 1, \ldots, N \tag{2.1}$$

Here $\xi_i(t)$ is an *intrinsic* noise term from the stochasticity of reaction events and

$$\begin{aligned}\Phi_i(\boldsymbol{x}(t)) = &\sum_{j,l,j\neq l} \left(k^-_{l,ij} x_i(t) - k^+_{ij,l} x_i(t) x_j(t)\right) + \frac{1}{2}\sum_{j,l,j\neq l}\left(k^+_{jl,i} x_j(t) x_l(t) - k^-_{i,jl} x_i(t)\right) \\ &+ \sum_j \left(\lambda_{ji} x_j(t) - \lambda_{ij} x_i(t)\right) + \sum_l \left(2k^-_{l,ii} x_l(t) - k^+_{ii,l} x_i(t) x_i(t)\right) + \sum_j \left(\frac{1}{2} k^+_{jj,i} x_j(t) x_j(t) - k^-_{i,jj} x_i(t)\right)\end{aligned} \tag{2.2}$$

These dynamical equations capture the formation of complex l from proteins i and j (with rate constant $k^+_{ij,l}$) and its dissociation ($k^-_{l,ij}$) and similar reactions where i itself is the complex (rate constants $k^+_{jl,i}$ and $k^-_{i,jl}$), as well as the unary change of species i into species j (λ_{ij}), e.g. by phosphorylation. We allow for the formation of both heterodimers and, with the last two terms in (2.2), homodimers. $\Phi_i(\boldsymbol{x}(t))$ encodes the deterministic part of Langevin dynamics (2.1) and can be written as $\Phi_i(\boldsymbol{x}(t)) = [\boldsymbol{S}\boldsymbol{f}(\boldsymbol{x}(t))]_i$ where $\boldsymbol{S}$ denotes the stoichiometric matrix with entries $0, \pm 1, \pm 2$ and $\boldsymbol{f}$ the vector of reaction fluxes.

The CLE-description is equivalent to the Chemical Fokker Planck equation that is obtained by truncating the Kramers-Moyal expansion of the underlying Master Equation after the first order in the inverse of the reaction volume V [33]. By construction of the CLE [3], the noise is Gaussian distributed with zero mean. Its covariance matrix

$$\langle \xi_i(t)\xi_j(t')\rangle = \boldsymbol{\Sigma}_{ij}(\boldsymbol{x})\delta(t-t') \tag{2.3}$$

shows it is uncorrelated in time (white noise) but multiplicative (with $\boldsymbol{x}$-dependent correlations, interpreted in the Itô convention [34]). One can show [35] that

$$\boldsymbol{\Sigma}(\boldsymbol{x}(t)) = \frac{\epsilon}{2}\, \boldsymbol{S}\, \mathrm{diag}(\boldsymbol{f}(\boldsymbol{x}(t)))\, \boldsymbol{S}^T \tag{2.4}$$

The inverse reaction volume $\epsilon = 1/V$ sets the amplitude of fluctuations, with the noise contribution diminishing as the reaction volume V (typically the cell volume) and hence the number of molecules increases. Following the approach in [36], we can study the time-dependent statistics of protein concentrations by appeal to the Martin–Siggia–Rose–Janssen–De Dominicis (MSRJD) formalism [37–39] (see also [40, 41] for

a systematic explanation of this formalism). We discretize time in small steps Δ and we denote by $P(\boldsymbol{x})$ the normalized probability distribution of protein concentration paths, i.e. the set of temporal trajectories $\boldsymbol{x} = \{x_i(t)\}$ satisfying (2.1); it can be deduced from $P(\boldsymbol{\xi})$, the Gaussian distribution of the stochastic fluctuations $\boldsymbol{\xi} = \{\xi_i(t)\}$ by

$$P(\boldsymbol{x}) = P(\boldsymbol{\xi}(\boldsymbol{x}))J(\boldsymbol{x}) \tag{2.5}$$

where $J(\boldsymbol{x}) = |\partial\boldsymbol{\xi}/\partial\boldsymbol{x}|$ is the Jacobian of this transformation and $J(\boldsymbol{x}) \equiv 1$ with the Itô convention [34, 41]. The normalization property of (2.5) can be rewritten as

$$\begin{aligned} 1 \equiv Z &= \int \prod_{it} dx_i(t) d\xi_i(t) P(\xi_i(t)) \delta\big(x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t)) - \Delta\overline{\xi}_i(t)\big) = \\ &= \int \prod_{it} \frac{dx_i(t) d\hat{x}_i(t)}{2\pi} d\xi_i(t) P(\xi_i(t)) e^{\mathrm{i}\hat{x}_i(t)\left[x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t)) - \Delta\overline{\xi}_i(t)\right]} \end{aligned} \tag{2.6}$$

where we have first expressed $\boldsymbol{\xi}$ in terms of $\boldsymbol{x}$ by using $\delta-$functions to impose (2.1) and we have then represented these as Fourier integrals over a set of auxiliary variables $\hat{\boldsymbol{x}}$. The noise appears via its average over the time interval Δ, $\overline{\xi}_i(t) = \frac{1}{\Delta}\int_t^{t+\Delta} \xi_i(t')dt'$, so that $\langle \overline{\xi}_i(t)\overline{\xi}_j(t')\rangle = \Delta^{-1}\boldsymbol{\Sigma}_{ij}(\boldsymbol{x})\delta(t-t')$. The integration over this Gaussian noise can be performed by means of the Hubbard-Stratonovich identity

$$\int D\boldsymbol{\xi} P(\boldsymbol{\xi}) e^{\pm \mathrm{i}\Delta\hat{\boldsymbol{x}}\cdot\overline{\boldsymbol{\xi}}} = e^{-\frac{\Delta}{2}\hat{\boldsymbol{x}}^{\boldsymbol{T}}\boldsymbol{\Sigma}\hat{\boldsymbol{x}}} \tag{2.7}$$

where the measure should be understood as $D\boldsymbol{\xi} = \prod_{it} d\xi_i(t)$. This brings Z into the final form

$$Z = \int D\boldsymbol{x}\, D\hat{\boldsymbol{x}}\, e^{\mathcal{H}} \tag{2.8}$$

The notation $D\boldsymbol{x}\, D\hat{\boldsymbol{x}}$ is a shorthand for $\prod_{it} dx_i(t)d\hat{x}_i(t)/(2\pi)$ and represents a path integral, ranging across all possible trajectories for $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. Each trajectory $\{x_i(t), \hat{x}_i(t)\}$ is weighted by a factor $\exp(\mathcal{H})$ with the action $\mathcal{H}$

$$\mathcal{H} = \sum_{it} \mathrm{i}\hat{x}_i(t)\left[x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t))\right] + \frac{\Delta}{2}\sum_{ijt} \mathrm{i}\hat{x}_i(t)\Sigma_{ij}(\boldsymbol{x}(t))\mathrm{i}\hat{x}_j(t) \tag{2.9}$$

Note that $Z \equiv 1$ by construction, so $\exp(\mathcal{H})$ defines a distribution (or more precisely a complex-valued measure) over trajectories, including the ones of the auxiliary variable $\hat{\boldsymbol{x}}$. This distribution will be the starting point for our model reduction strategy, which aims to derive an accurate description of the dynamics of a chosen *subnetwork* that accounts also for the dynamical effects of the embedding into its environment, the *bulk*.

2.1. Projection Methods

As explained in the introduction, model reduction to subnetwork dynamics can also be achieved using projection methods [27, 42]. Here one defines an operator that projects the dynamics onto observables describing a set of variables of interest. These relevant variables can be slow degrees of freedom, or in the context of optimal prediction methods [23, 24, 43] those variables that are better resolved by measurement. The contributions of the remaining, less relevant variables, are then isolated into a *memory* kernel and a *random force*.

Briefly, if $x_i(t)$ is a subnetwork concentration, with $i = 1, \ldots, N^{\mathrm{s}}$ where N^{s} is the size of the subnetwork, the projection approach yields a dynamical equation of the general form (see e.g. [42])

$$\frac{\partial x_i(t)}{\partial t} = \sum_{j=1}^{N^{\mathrm{s}}} x_j(t) \underbrace{\Omega_{ji}}_{\text{Rate matrix}} + \sum_{j=1}^{N^{\mathrm{s}}} \int_0^t dt' x_j(t') \underbrace{M_{ji}(t-t')}_{\text{Memory function}} + \underbrace{r_i(t)}_{\text{Random force}} \tag{2.10}$$

The rate matrix Ω_{ji} in the first term on the r.h.s. describes processes that are local in time and involving solely subnetwork observables. The second term, the memory, captures the effects that past subnetwork states at

time t' have on the dynamics at time t. The strength of these effects is given by the memory function, which depends on the time lag $t - t'$ as well as the species involved. The intuition behind the memory is that the subnetwork state at t' affects the bulk dynamics, which in turn feeds back into the subnetwork dynamics at t. The final term, the random force, represents the effect of unknown bulk initial conditions, hence a form of extrinsic noise, and is distinct from any stochasticity of the original dynamics (intrinsic noise). Note that the projected dynamics formally describes the *conditional average* of the subnetwork concentrations, where the conditioning is on the initial values. Fluctuations due to intrinsic noise are therefore not represented explicitly. The presence of memory and the random force – which has non-trivial temporal correlations – make the projected subnetwork equations non-Markovian, as expected on general grounds given that model reduction to a subnetwork is a form of coarse-graining. Including these non-Markovian effects means less information is lost, giving a more accurate prediction of time courses than any Markovian approximation [26, 43]. We note finally that by choosing more general observables, e.g. products of concentration fluctuations, a nonlinear analogue of (2.10) can be constructed and it is this form that we compare our framework to in Sec. 5.3 and Appendix E.

3. Gaussian Variational Approximation

Returning to our probabilistic setup, our model reduction task is in principle to take the full distribution over network trajectories in (2.8) and marginalize it, i.e. integrate over all possible trajectories of the bulk variables. This is clearly intractable, however, while it would be straightforward if the distribution were Gaussian (see Sec. 4 below). We therefore now construct a Variational Gaussian Approximation (GVA) to the network trajectory distribution. Note that in (2.8) we implicitly assumed the initial concentrations $\boldsymbol{x}(0)$ to be fixed; this can be relaxed by multiplying by the relevant distribution $P_0(\boldsymbol{x}(0))$

$$P(\boldsymbol{x}, \hat{\boldsymbol{x}}) = e^{\mathcal{H}(\boldsymbol{x}, \hat{\boldsymbol{x}})} P_0(\boldsymbol{x}(0)) \tag{3.1}$$

We approximate P by a variational distribution $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ of Gaussian form

$$P(\boldsymbol{x}, \hat{\boldsymbol{x}}) \approx Q(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \mathcal{N}(\boldsymbol{x}, \hat{\boldsymbol{x}} | \boldsymbol{\mu}_{\text{gen}}, \boldsymbol{C}_{\text{gen}}) \tag{3.2}$$

This Gaussian is completely determined by the vector of mean values $\boldsymbol{\mu}_{\text{gen}}$ and the covariance matrix $\boldsymbol{C}_{\text{gen}}$, where the "gen" subscript indicates "generalized" moments including the auxiliary variables $\hat{\boldsymbol{x}}(t)$

$$\boldsymbol{\mu}_{\text{gen}}(t) = \begin{pmatrix} \langle \boldsymbol{x}(t) \rangle_Q \\ -\mathrm{i} \langle \hat{\boldsymbol{x}}(t) \rangle_Q \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}(t) \\ -\mathrm{i} \hat{\boldsymbol{\mu}}(t) \end{pmatrix} \tag{3.3}$$

$$\boldsymbol{C}_{\text{gen}}(t, t') = \left\langle \begin{pmatrix} \delta \boldsymbol{x}(t) \\ -\mathrm{i} \delta \hat{\boldsymbol{x}}(t) \end{pmatrix} \begin{pmatrix} \delta \boldsymbol{x}(t') & -\mathrm{i} \delta \hat{\boldsymbol{x}}(t') \end{pmatrix} \right\rangle_Q = \begin{pmatrix} \boldsymbol{C}(t, t') & \boldsymbol{R}(t, t') \\ \boldsymbol{R}(t', t)^T & \boldsymbol{B}(t, t') \end{pmatrix} \tag{3.4}$$

with $\delta \boldsymbol{x}(t) = \boldsymbol{x}(t) - \boldsymbol{\mu}(t)$ and $\delta \hat{\boldsymbol{x}}(t) = \hat{\boldsymbol{x}}(t) - \hat{\boldsymbol{\mu}}(t)$. For the sake of brevity, we will write $\langle \rangle_Q$ as simply $\langle \rangle$ and the notation used has the following meaning

$$\begin{aligned} \boldsymbol{C}(t, t') &= \langle \delta \boldsymbol{x}(t) \delta \boldsymbol{x}^T(t') \rangle = \langle \boldsymbol{x}(t) \boldsymbol{x}^T(t') \rangle - \boldsymbol{\mu}(t) \boldsymbol{\mu}^T(t') \\ \boldsymbol{R}(t, t') &= -\mathrm{i} \langle \delta \boldsymbol{x}(t) \delta \hat{\boldsymbol{x}}^T(t') \rangle = -\mathrm{i} \langle \boldsymbol{x}(t) \hat{\boldsymbol{x}}^T(t') \rangle + \mathrm{i} \boldsymbol{\mu}(t) \hat{\boldsymbol{\mu}}^T(t') \\ \boldsymbol{B}(t, t') &= -\langle \delta \hat{\boldsymbol{x}}(t) \delta \hat{\boldsymbol{x}}^T(t') \rangle = -\langle \hat{\boldsymbol{x}}(t) \hat{\boldsymbol{x}}^T(t') \rangle + \hat{\boldsymbol{\mu}}(t) \hat{\boldsymbol{\mu}}^T(t') \end{aligned} \tag{3.5}$$

From general results for MSRJD path integrals [40, 44] it follows that $R_{ij}(t, t')$ has the meaning of a local response of $x_i(t)$ to a perturbing field $-\mathrm{i}\hat{x}_j(t')$ applied to some other species j at an earlier time t'; it is therefore non-vanishing only for $t > t'$. Also, $B_{ij}(t, t')$ is expected to vanish for all times t and t', as is $\hat{\mu}_i(t)$: we refer to [45] for a derivation of this property from the fact that $Z \equiv 1$.

The optimal Gaussian approximation can be found from the stationary points of the Kullback-Leibler (KL) divergence between $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ and $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$ [12], i.e. from the conditions

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{\mu}_{\text{gen}}} = 0 \qquad \frac{\partial \mathrm{KL}}{\partial \boldsymbol{C}_{\text{gen}}} = 0 \tag{3.6}$$

These conditions give differential equations for the time-dependent parameters of the optimal Gaussian $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$, i.e. its first two moments. The calculation (see Appendix A) is essentially a reformulation in the MSRJD formalism of the variational derivation provided by Archambeau et al. [15], except that by focussing directly on the stationarity conditions of the KL-divergence we can deal with the concentration dependence of the diffusion matrix $\boldsymbol{\Sigma}(\boldsymbol{x}(t))$. We note that, as $P(\boldsymbol{x}, \hat{\boldsymbol{x}})$ and $Q(\boldsymbol{x}, \hat{\boldsymbol{x}})$ are complex-valued, stationarity of the KL-divergence is only a *heuristic* strategy for finding the optimal approximation, but one that is validated *a posteriori* by comparison with the results of [15].

The time evolution of the Gaussian Variational Approximation means is given by

$$\frac{d\mu_i(t)}{dt} = \langle \Phi_i(\boldsymbol{x}(t)) \rangle \tag{3.7}$$

so that $\boldsymbol{\mu}(t)$ evolves with the average drift of the initial equations of motion (2.1). The average is over the instantaneous Gaussian distribution with mean $\boldsymbol{\mu}(t)$ and covariance $\boldsymbol{C}(t,t)$, so that nonlinear terms in $\Phi_i(\boldsymbol{x})$ retain information about the fluctuations in the dynamics, unlike the Linear Noise Approximation (see below).

In the unapproximated dynamics, the auxiliary variables $\hat{x}_i(t)$ have vanishing moments of any order, and (within the Itô discretization [34, 41]) equal-time responses. It turns out that the GVA preserves these properties, giving $\hat{\boldsymbol{\mu}}(t) = \boldsymbol{B}(t,t') = 0$ and $\boldsymbol{R}(t,t) = 0$.

It remains to discuss the predictions of the GVA for the correlations $\boldsymbol{C}(t,t')$ and the responses $\boldsymbol{R}(t,t')$ for $t > t'$. One finds that these are what would be predicted by linearization of the original Langevin dynamics (2.1) around the time-dependent means $\boldsymbol{\mu}(t)$

$$\frac{d(\boldsymbol{x}(t) - \boldsymbol{\mu}(t))}{dt} = \boldsymbol{K}(t)(\boldsymbol{x}(t) - \boldsymbol{\mu}(t)) + \boldsymbol{\xi}(t) \tag{3.8}$$

with an effective rate matrix $\boldsymbol{K}(t)$ defined as

$$K_{ij}(t) = \left\langle \frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} \right\rangle \tag{3.9}$$

and white Gaussian additive noise $\boldsymbol{\xi}(t)$ with

$$\langle \boldsymbol{\xi}(t) \boldsymbol{\xi}^T(t') \rangle = \langle \boldsymbol{\Sigma}(\boldsymbol{x}(t)) \rangle \delta(t - t') \tag{3.10}$$

Thus, the noise covariance of the GVA is the true one averaged over the approximating Gaussian distribution. The result (3.9) coincides with that of Archambeau et al. [15], while (3.10) allows one to generalize their approach to the multiplicative noise case in which the GVA cannot match the true noise covariance instantaneously. The GVA responses and correlations are the ones that follow straightforwardly from (3.8) (see Appendix A).

We stress that the GVA differs from the Linear Noise Approximation (LNA), which is obtained within the van Kampen system size expansion of the Master Equation [3]. The LNA also produces a Gaussian approximation, but linearizes around the macroscopic (deterministic) trajectory (see [35]) given by

$$\frac{d\mu_i(t)}{dt} = \Phi_i(\boldsymbol{\mu}(t)) \tag{3.11}$$

The evolution of the means is therefore completely decoupled from any fluctuation effects. In the GVA, on the other hand, the mean dynamics (3.7) is affected by the equal-time fluctuations $\boldsymbol{C}(t,t)$, as are the effective rate matrix (3.9) and the noise covariance (3.10). GVA and LNA only become equivalent when these fluctuations vanish, i.e. in the limit of large reaction volumes.

4. Subnetwork dynamics within GVA

We can now exploit the GVA results to perform the marginalization over the bulk dynamics and thus obtain a reduced description of the subnetwork dynamics that still accounts implicitly for the presence of the bulk. It is here that we first go beyond existing work.

We separate the concentrations and auxiliary variables (x_i and $\hat{x}_i$) into subnetwork and bulk. It is easiest to assume they are numbered so that $i = 1, \ldots, N^{\mathrm{s}}$ labels subnetwork species while the remaining $N^{\mathrm{b}} = N - N^{\mathrm{s}}$ species $i = N^{\mathrm{s}}, \ldots, N$ are in the bulk. This gives a block structure for $\boldsymbol{C}$ and $\boldsymbol{R}$

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}^{\mathrm{ss}} & \boldsymbol{C}^{\mathrm{sb}} \\ \boldsymbol{C}^{\mathrm{bs}} & \boldsymbol{C}^{\mathrm{bb}} \end{pmatrix}$$

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}^{\mathrm{ss}} & \boldsymbol{R}^{\mathrm{sb}} \\ \boldsymbol{R}^{\mathrm{bs}} & \boldsymbol{R}^{\mathrm{bb}} \end{pmatrix}$$

where the superscripts (s or b) refer to subnetwork or bulk respectively. To retain bulk effects on the subnetwork dynamics the off-diagonal blocks (sb and bs) are clearly essential. Indeed, if one constructs a restricted GVA where these blocks are constrained to vanish, corresponding to a factorized Gaussian approximation, then no memory terms appear in the reduced subnetwork dynamics.

Bulk marginalization is now conceptually straightforward: marginals of a joint Gaussian are still Gaussian, with the appropriate mean and (in our case, generalized) covariance. The marginal distribution over subnetwork trajectories $\{\boldsymbol{x}^{\mathrm{s}}(t), \hat{\boldsymbol{x}}^{\mathrm{s}}(t)\}$ is therefore Gaussian with second moments $\boldsymbol{C}^{\mathrm{ss}}(t,t')$, $\boldsymbol{R}^{\mathrm{ss}}(t,t')$ and $\boldsymbol{B}^{\mathrm{ss}}(t,t') \equiv 0$. To write down the distribution itself, one just has to find the inverse of this large (generalized) covariance matrix, as this is what features in the exponent of the Gaussian. Finally one has to read off *what dynamical equations this Gaussian distribution over subnetwork trajectories encodes.*

An equivalent and technically simpler route is to write down the dynamics (3.8) implied by the GVA for the full network and eliminate the bulk degrees of freedom explicitly. We write the subnetwork and bulk parts of $\delta\boldsymbol{x}(t) = \boldsymbol{x}(t) - \boldsymbol{\mu}(t)$ as $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ and $\delta\boldsymbol{x}^{\mathrm{b}}(t)$, respectively. Similarly we decompose the rate matrix in equation (3.8) as

$$\boldsymbol{K}(t) = \begin{pmatrix} \boldsymbol{K}^{\mathrm{ss}}(t) & \boldsymbol{K}^{\mathrm{sb}}(t) \\ \boldsymbol{K}^{\mathrm{bs}}(t) & \boldsymbol{K}^{\mathrm{bb}}(t) \end{pmatrix}$$

Then (3.8) decomposes into

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}}{dt} = \boldsymbol{K}^{\mathrm{ss}}(t)\delta\boldsymbol{x}^{\mathrm{s}} + \boldsymbol{K}^{\mathrm{sb}}(t)\delta\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{\xi}^{\mathrm{s}} \tag{4.1}$$

$$\frac{d\delta\boldsymbol{x}^{\mathrm{b}}}{dt} = \boldsymbol{K}^{\mathrm{bb}}(t)\delta\boldsymbol{x}^{\mathrm{b}} + \boldsymbol{K}^{\mathrm{bs}}(t)\delta\boldsymbol{x}^{\mathrm{s}} + \boldsymbol{\xi}^{\mathrm{b}} \tag{4.2}$$

One can now eliminate $\delta\boldsymbol{x}^{\mathrm{b}}$ explicitly as

$$\delta\boldsymbol{x}^{\mathrm{b}}(t) = \mathrm{T}\left[e^{\int_0^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\delta\boldsymbol{x}^{\mathrm{b}}(0) + \int_0^t \mathrm{T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\left(\boldsymbol{K}^{\mathrm{bs}}(t')\delta\boldsymbol{x}^{\mathrm{s}}(t') + \boldsymbol{\xi}^{\mathrm{b}}(t')\right)dt' \tag{4.3}$$

where $\delta\boldsymbol{x}^{\mathrm{b}}(0)$ is the bulk initial condition and the T indicates that the exponentials are time-ordered, with factors of $\boldsymbol{K}^{\mathrm{bb}}(s)$ at each order of the exponential expansion arranged so that time increases from right to left. Substituting into (4.1) gives the reduced subnetwork dynamics in the form

$$\frac{d\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} = \boldsymbol{K}^{\mathrm{ss}}(t)\delta\boldsymbol{x}^{\mathrm{s}}(t) + \int_0^t \boldsymbol{M}^{\mathrm{ss}}(t,t')\delta\boldsymbol{x}^{\mathrm{s}}(t') + \boldsymbol{\chi}(t) \tag{4.4}$$

All subnetwork reactions are retained directly in the local-in-time term $\boldsymbol{K}^{\mathrm{ss}}$ while the bulk gives a contribution via the additional non-local-in-time terms. The second term in particular contains the memory function

$$\boldsymbol{M}^{\mathrm{ss}}(t,t') = \boldsymbol{K}^{\mathrm{sb}}(t)\,\mathrm{T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\mathrm{bb}}(s)ds}\right]\boldsymbol{K}^{\mathrm{bs}}(t') \tag{4.5}$$

This memory function has a *boundary* structure [26]: memory terms only appear in the reduced dynamical equations for "boundary species", defined as those subnetwork species that interact with the bulk and that therefore have at least some nonzero coefficients in the relevant row of $\boldsymbol{K}^{\mathrm{sb}}$. We defer discussion of the effective noise $\boldsymbol{\chi}$ in (4.4) to Appendix B; its covariance is given in (B.5).

The expression (4.5) for the memory function is appealing conceptually – signals enter the bulk at time t', propagate around the bulk and then return to the subnetwork at time t – and the applicability to time-dependent effective rate matrices $\boldsymbol{K}^{\mathrm{bb}}(t)$ makes it more general than e.g. projection approaches [26]. For computational purposes a time-dependent $\boldsymbol{K}^{\mathrm{bb}}(t)$ is inconvenient, however, because of the need to evaluate time-ordered exponentials. From now on we therefore consider dynamics around a *steady state*, where the effective rate and diffusion matrices become constant in time, $\boldsymbol{K}(t) \equiv \boldsymbol{K}$ and $\langle \boldsymbol{\Sigma}(\boldsymbol{x}(t)) \rangle \equiv \boldsymbol{\Sigma}$. In this simpler case (4.5) becomes

$$\boldsymbol{M}^{\mathrm{ss}}(t,t') = \boldsymbol{K}^{\mathrm{sb}} e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} \boldsymbol{K}^{\mathrm{bs}} \tag{4.6}$$

This then agrees fully with the memory function in the reduced description derived via projection methods [26], as summarized in Appendix E.1. Given this agreement we can refer to [26] for a systematic analysis of typical amplitudes and timescales of the memory, which encode its overall effect on dynamics, in the case of a mass action kinetics such as (2.1).

In contrast to the memory function, the effective noise term (see (B.7) in Appendix B) in our marginalization approach differs from its apparent analogue in the projection method, the random force. This is because, as explained in Sec. 2.1, the projection approach works with conditionally averaged observables and hence eliminates (in a linear theory such as the GVA) all intrinsic noise contributions. The random force thus contains only the extrinsic noise from the initial uncertainty about the bulk state (see Appendix E). Our effective noise, on the other hand, contains in addition the implicit noise effects.

So far no restrictions have been placed on the relative timescales of bulk and subnetwork. If we specialize to the case of a slowly varying subnetwork embedded in a fast bulk, equation (4.4) can be shown (see [46]) to reduce to the slow-scale LNA introduced by Thomas et al. [25, 47], and this is a useful consistency check for our approach.

5. Nonlinear corrections by perturbation theory

5.1. Nonlinear memory

As explained, our general approach is to derive the reduced dynamics of a subnetwork by marginalizing out bulk degrees of freedom to obtain an "effective" action that involves *solely* subnetwork variables

$$\mathcal{H}_{\mathrm{eff}}(\boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}) = \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{\mathcal{H}(\boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}})} \tag{5.1}$$

where $D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}}$ indicates the integration over bulk trajectories and $\mathcal{H}$ is the dynamical action (2.9) of the whole system. So far we have done this within a Gaussian approximation for $\mathcal{H}$, where the marginalization can be done in closed form; for ease of derivation we sidestepped the dynamical action in this case and eliminated the bulk variables directly from the corresponding dynamical equations.

To go beyond the resulting simple linear expression for the memory function (Sec. 4) we now use a perturbative expansion to capture nonlinear contributions in the dynamics, specifically cubic terms in the effective action. We decompose the action $\mathcal{H}$ into a "non-interacting" part $\mathcal{H}_0$ containing only the purely Gaussian terms, i.e. those that are quadratic in all variables $\delta\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, and an interacting piece $\Delta\mathcal{H}$ containing higher powers. Assuming formally that the rate constants defining $\Delta\mathcal{H}$ are small, we can expand the effective action (5.1) as

$$\begin{aligned} \mathcal{H}_{\mathrm{eff}} &= \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{(\mathcal{H}_0 + \Delta\mathcal{H})} \\ &= \ln \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} e^{\mathcal{H}_0} \left(1 + \int D\boldsymbol{x}^{\mathrm{b}} D\hat{\boldsymbol{x}}^{\mathrm{b}} Q_0(\boldsymbol{x}^{\mathrm{b}}, \hat{\boldsymbol{x}}^{\mathrm{b}} | \boldsymbol{x}^{\mathrm{s}}, \hat{\boldsymbol{x}}^{\mathrm{s}}) \Delta\mathcal{H} + \mathcal{O}(\Delta\mathcal{H}^2) + \dots \right) \end{aligned} \tag{5.2}$$

Here

$$\begin{aligned} -\Delta\mathcal{H} = \int_0^T dt \Big\{ \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}} \Big[&\boldsymbol{K}^{\mathrm{s,ss}}\big(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}\big) + \boldsymbol{K}^{\mathrm{s,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\big) + \boldsymbol{K}^{\mathrm{s,bb}}\big(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\big) + \\ &\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}} \Big[\boldsymbol{K}^{\mathrm{b,ss}}\big(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}\big) + \boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\big) + \boldsymbol{K}^{\mathrm{b,bb}}\big(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\big) \Big] \Big\} \end{aligned} \tag{5.3}$$

collects all *quadratic* terms in $\delta\boldsymbol{x}$ in the original mass action dynamics (2.2): thus the perturbative approach we develop here is targeted at this type nonlinearities, i.e. binary reactions. For simplicity we have dropped terms arising from the concentration dependence of the noise covariance here; see Appendix C. We have written directly the continuous time ($\Delta \to 0$) version and introduced an explicit final time T for trajectories. We have also introduced convenient shorthands to separate terms that depend on different combinations of subnetwork or bulk variables. We use the $\boldsymbol{a} \circ \boldsymbol{b}$ notation, not to be confused with a Hadamard (elementwise) product, to denote the outer product $\boldsymbol{a}\boldsymbol{b}^T$ rearranged into a single (column) vector. This vector then has as its entries all possible componentwise products $a_i b_j$ so that e.g. $\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}$ is a vector of dimension $N^{\mathrm{s}}N^{\mathrm{b}}$. Where the two vectors are from the same index range (both "s" or both "b") we use the ordered products only: $\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}$ is the vector containing the $N^{\mathrm{s}}(N^{\mathrm{s}}+1)/2$ products $\delta x_i \delta x_j$ with $1 \le i \le j \le N^{\mathrm{s}}$. The matrices $\boldsymbol{K}^{\mathrm{s,ss}}$ etc. are then defined to contain the appropriate coefficients to ensure that, for example,

$$\left[\boldsymbol{K}^{\mathrm{s,ss}}\left(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}\right)\right]_i = \sum_{j,l=1,j\neq l}^{N^{\mathrm{s}}} \left(\frac{1}{2}k^+_{jl,i}\delta x_j \delta x_l - k^+_{ij,l}\delta x_i \delta x_j\right) + \frac{1}{2}\sum_{j=1}^{N^{\mathrm{s}}} k^+_{jj,i}\delta x_j \delta x_j - \sum_{l=1}^{N^{\mathrm{s}}} k^+_{ii,l}\delta x_i \delta x_i \tag{5.4}$$

Because of the ordering of the $\circ$ product this requires that the elements of $\boldsymbol{K}^{\mathrm{s,ss}}$ are defined as $K_{i,jl} = k^+_{jl,i} - (\delta_{ij} + \delta_{il})\sum_{m=1}^{N^{\mathrm{s}}} k^+_{jl,m}$ for $j < l$ and $K_{i,jj} = \frac{1}{2}k^+_{jj,i} - \delta_{ij}\sum_{m=1}^{N^{\mathrm{s}}} k^+_{ii,m}$. These coefficients define the terms in the drift (2.2) for subnetwork variables that depend on quadratic combinations of subnetwork variables. $\boldsymbol{K}^{\mathrm{s,sb}}$, $\boldsymbol{K}^{\mathrm{b,sb}}$, $\boldsymbol{K}^{\mathrm{b,bb}}$ are used as shorthands for sets of rate constants involved, respectively, in the reaction of a bulk and subnetwork species to give a complex in the subnetwork ($\boldsymbol{K}^{\mathrm{s,sb}}$) or in the bulk ($\boldsymbol{K}^{\mathrm{b,sb}}$) or two bulk species ($\boldsymbol{K}^{\mathrm{b,bb}}$) to give another bulk species. We will set $\boldsymbol{K}^{\mathrm{s,bb}} = \boldsymbol{K}^{\mathrm{b,ss}} \equiv 0$, i.e. we do not include in the subnetwork complexes formed by 2 bulk species (and viceversa we do not include in the bulk complexes formed by 2 subnetwork species). This is in line with the spirit of our model reduction setting where subnetwork species are classified as the well-characterized ones (e.g. because they have been fully resolved biochemically or can be monitored experimentally with accuracy) while bulk species are assumed to be less characterized.

As (5.2) shows, the effective action is calculated by averaging $\Delta\mathcal{H}$ over the distribution of the bulk variables *conditional* on the subnetwork variables. This average is governed by the Gaussian baseline distribution $Q_0 = \exp(\mathcal{H}_0)$, so that also the conditional statistics are Gaussian (see Appendix C). We could choose Q_0 as the steady state GVA distribution, $Q_0 = Q$, but this would lead to terms *linear* in $\hat{\boldsymbol{x}}$ in $\mathcal{H}_0$ that could be interpreted as effective fields. These generate nonzero means, which are inconvenient for the conditional Gaussian averaging. To see how such terms arise we write a schematic version of the drift part of the action

$$\int_0^T dt \sum_i \mathrm{i}\hat{x}_i\, \Phi_i(\boldsymbol{x}) = \int_0^T dt \sum_i \mathrm{i}\hat{x}_i\, \Phi_i(\boldsymbol{\mu} + \delta\boldsymbol{x}) = \int_0^T dt \sum_i \mathrm{i}\hat{x}_i\, \left[\Phi_i(\boldsymbol{\mu}) + \mathcal{O}(\delta\boldsymbol{x}) + \mathcal{O}(\delta\boldsymbol{x}^2)\right] \tag{5.5}$$

where in the last step we have expanded the drift Φ_i around the $\boldsymbol{\mu}$, up to quadratic terms in $\delta\boldsymbol{x}$. Linear terms appear if we use $\boldsymbol{\mu}$ as obtained from the GVA since then $\Phi_i(\boldsymbol{\mu})$ is not zero. It is therefore more convenient to use the (steady state) LNA, see also Sec. 2 and [35], for which $\Phi_i(\boldsymbol{\mu}) = 0$. The interpretation of $\boldsymbol{\mu}$ is then not as a *mean* steady state concentration in a system with intrinsic noise, but rather as the steady state of the corresponding *deterministic* dynamics. The resulting quadratic action is

$$\begin{aligned}\mathcal{H}_0 = \int_0^T dt \Big\{ &\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\left[\partial_t \delta\boldsymbol{x}^{\mathrm{s}} - \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{sb}}\delta\boldsymbol{x}^{\mathrm{b}}(t)\right] + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{ss}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{sb}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) \\ &\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,T}(t)\left[\partial_t \delta\boldsymbol{x}^{\mathrm{b}} - \boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t) - \boldsymbol{K}^{\mathrm{bb}}\delta\boldsymbol{x}^{\mathrm{b}}(t)\right] + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{bb}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}}(t) + \frac{1}{2}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}\,\mathrm{T}}(t)\boldsymbol{\Sigma}^{\mathrm{bs}}\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)\Big\}\end{aligned} \tag{5.6}$$

The conditional statistics of the bulk variables given the subnetwork ones, which we need for our perturbation expansion, can be worked out from (5.6) using general results for Gaussian conditioning (see e.g. [11]). The conditional mean of the bulk auxiliary variables is

$$\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) = \int_t^T dt' e^{-(\boldsymbol{K}^{\mathrm{bb}})^T(t-t')}(\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{5.7}$$

where we use the superscript b|s (and bb|s below) to indicate bulk averages conditioned on the entire subnetwork trajectories. This can be viewed as the solution of the differential equation

$$\frac{d\,\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t)}{dt} = -(\boldsymbol{K}^{\mathrm{bb}})^T \mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) + (\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) \tag{5.8}$$

backwards in time, starting from the final boundary condition $\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(T) = 0$. This is consistent with the theory of conditional Markov processes, for which calculating conditional distributions requires information to propagate both forward and backward in time (see also [48]). Note that unlike the marginal auxiliary means, the conditional auxiliary means can be nonzero as long as the subnetwork conditional variables we condition on are also nonzero.

The conditional means of the bulk concentrations follow from (5.6) and differ from the marginal means by $\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\mu}^{\mathrm{b|s}}(t) - \boldsymbol{\mu}^{\mathrm{b}} = \boldsymbol{\nu}(t) + \hat{\boldsymbol{\nu}}(t)$ where

$$\boldsymbol{\nu}(t) = \int_0^t dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} \boldsymbol{K}^{\mathrm{bs}} \delta\boldsymbol{x}^{\mathrm{s}}(t') \qquad \hat{\boldsymbol{\nu}}(t) = -\int_0^T dt' \boldsymbol{C}^{\mathrm{bb|s}}(t,t')(\boldsymbol{K}^{\mathrm{sb}})^T \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{5.9}$$

Here $\boldsymbol{\nu}(t)$ is the deterministic part of the conditional time evolution of the bulk concentrations, while $\hat{\boldsymbol{\nu}}(t)$ involves $\boldsymbol{C}^{\mathrm{bb|s}}(t,t')$, the conditional correlator of the bulk concentration fluctuations, therefore can be interpreted as carrying the "stochastic" contributions from the bulk, i.e. the uncertainty about its initial values and its intrinsic noise.

To complete the description of the conditional bulk statistics one needs the conditional second moments. These can be obtained by setting the subnetwork variables in (5.6) to zero (see [46]). As such, the conditional correlator of auxiliary variables $\boldsymbol{B}^{\mathrm{bb|s}}(t,t') = 0$ and the equal-time conditional response $\boldsymbol{R}^{\mathrm{bb|s}}(t,t) \equiv 0$ both vanish, as in the full dynamics. (Causal responses $\boldsymbol{R}^{\mathrm{bb|s}}(t,t')$ are nonzero, but do not appear in our first order perturbation theory.) $\boldsymbol{C}^{\mathrm{bb|s}}(t,t')$, which also appears in the expression (5.9), is given by

$$\boldsymbol{C}^{\mathrm{bb|s}}(t,t'') = \int_0^{\min(t,t'')} dt' e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')} \boldsymbol{\Sigma}^{\mathrm{bb}} e^{(\boldsymbol{K}^{\mathrm{bb}})^T(t''-t')} + e^{\boldsymbol{K}^{\mathrm{bb}}t} \boldsymbol{C}^{\mathrm{bb}}(0,0) e^{(\boldsymbol{K}^{\mathrm{bb}})^T t''} \tag{5.10}$$

As generally the case in Gaussian conditioning, the second order statistics do not depend on the values of the observables (here: the subnetwork concentrations) being conditioned on.

We can now find the reduced subnetwork action from (5.2) by performing the Gaussian averages over the bulk variables with the conditional statistics as given by (5.7), (5.9) and (5.10). From this reduced action one can then read off the nonlinear reduced dynamics of the subnetwork. We leave the details for Appendix C and state directly the result

$$\begin{aligned}\frac{d\,\delta\boldsymbol{x}^{\mathrm{s}}(t)}{dt} &= \boldsymbol{K}^{\mathrm{ss}}\delta\boldsymbol{x}^{\mathrm{s}}(t) + \boldsymbol{K}^{\mathrm{s,ss}}(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t)) + \int_0^t dt' \boldsymbol{M}^{\mathrm{ss}}(t,t')\delta\boldsymbol{x}^{\mathrm{s}}(t')dt' \\ &\quad + \int_0^t dt' \int_0^{t'} dt'' \boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big) + \boldsymbol{\chi}(t)\end{aligned} \tag{5.11}$$

where in $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')$ the times are ordered, $t > t' > t''$. The coloured noise $\boldsymbol{\chi}(t)$ has covariance $\langle\boldsymbol{\chi}(t)\boldsymbol{\chi}(t')^T\rangle = \boldsymbol{N}_0^{\mathrm{ss}}(t,t') + \boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ consisting of a linear and nonlinear contribution. While $\boldsymbol{N}_0^{\mathrm{ss}}(t,t')$ is the covariance of a Gaussian coloured noise (namely (B.7)), $\boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ is multiplicative as it is linearly dependent on $\boldsymbol{x}^{\mathrm{s}}$.

We see that reactions within the subnetwork contribute to the reduced dynamical equations only via $\boldsymbol{K}^{\mathrm{ss}}$ and $\boldsymbol{K}^{\mathrm{s,ss}}$ (a compact notation for the rate constants of respectively linear and nonlinear couplings inside the subnetwork): subnetwork reactions are fully reproduced in local-in-time terms, which is one of the desiderata of our reduced description.

While $\boldsymbol{M}^{\mathrm{ss}}(t,t')$ is the memory function given by (4.6), the nonlinear memory function, $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')$, can be read off from the coefficients of terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$ in the effective perturbative action (5.2), which themselves come from the $\boldsymbol{\nu}(t)$ part of the conditional mean (5.9). The effective noise covariance, $\boldsymbol{N}_1^{\mathrm{ss}}(t,t')$

is similarly extracted from terms $\sim \hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$ that arise from $\hat{\boldsymbol{\nu}}(t)$. We refer to Appendix C for details of the derivation and just state the final result (for $t > t' > t''$)

$$\begin{aligned}\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big) = \\ &+ \boldsymbol{K}^{\mathrm{s,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big)\delta(t'-t) \\ &+ \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big) \\ &+ 2\int_{t'}^{t} ds\, \boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-s)}\boldsymbol{K}^{\mathrm{b,bb}}\big(e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(s-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big)\end{aligned} \tag{5.12}$$

This form emphasizes that the nonlinear memory that results from our approach involves products of concentration fluctuations at *different* times t' and t'' in the past. In the projection approach, on the other hand, only equal-time products from the past appear [26]. While our result therefore looks more complicated conceptually, its numerical implementation is in fact more efficient than for the projected subnetwork equations as we discuss next.

5.2. Numerical implementation

We will denote by $\boldsymbol{\mathcal{M}}(t)$ the total memory appearing in the subnetwork dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$. This is given by the two integral terms in (5.11). For numerical purposes, each of these can be represented via the solutions of additional differential equations. Hence, one can solve the *integro-differential* equations (5.11) using only a *differential* equation solver but in an enlarged space of variables. To see this, we first note that $\boldsymbol{\mathcal{M}}(t)$ can be expressed in terms of the $\boldsymbol{\nu}$-part of the bulk conditional means as

$$\begin{aligned}\boldsymbol{\mathcal{M}}(t) = \boldsymbol{K}^{\mathrm{sb}}\boldsymbol{\nu}(t) + \boldsymbol{K}^{\mathrm{s,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{\nu}(t)\big) \\ + \int_0^t dt'\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\left[\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{\nu}(t')\big) + \boldsymbol{K}^{\mathrm{b,bb}}\big(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t')\big)\right]\end{aligned} \tag{5.13}$$

The bulk means (5.9) are the solution of the N^{b} differential equations

$$\frac{d\boldsymbol{\nu}}{dt} = \boldsymbol{K}^{\mathrm{bb}}\boldsymbol{\nu} + \boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}} \tag{5.14}$$

with $\boldsymbol{\nu}(0) = 0$. The $\boldsymbol{K}^{\mathrm{b,sb}}$ and $\boldsymbol{K}^{\mathrm{b,bb}}$ pieces of (5.13) contain an additional time integral; to translate them into differential equations for additional variables we need to apply a decomposition into eigenvalues and eigenvectors of $\boldsymbol{K}^{\mathrm{bb}}$, similarly to the procedure in [29]. This requires N^{b} auxiliary variables in addition to the $\boldsymbol{\nu}$, giving $2N^{\mathrm{b}}$ in total; see Appendix D for details.

It is notable here that for subnetwork equations derived by projection methods, a similar conversion into differential equations requires a number of additional variables growing *quadratically* with the number of bulk nodes [26]. The smaller, *linear*, number of auxiliary variables in our method is a significant computational advantage for the typical case where the bulk network is large, e.g. because the overall network itself is large while the subnetwork contains only a few, experimentally well resolved, species.

5.3. Comparison with projection methods

The comparison of our perturbatively corrected GVA approach to nonlinear projection methods is not as straightforward as for the linearized equations. The reason is that memory functions and the noise correlators derived in the two approaches are not the same if taken *separately*. But remarkably, we can show that the *combination* of the memory term and the coloured noise from equation (5.11) provides an approximation of the subnetwork reduced dynamics that is *equivalent* to the one obtained by projection techniques (Appendix E.2). This holds in the *limit of negligible intrinsic noise* and *up to quadratic order* in $\delta\boldsymbol{x}$. The proof of this equivalence is rather non-trivial, both conceptually and algebraically, so we defer it to Appendix E.4.

6. Application to biochemical networks

In this section we illustrate the GVA reduction method by applying it to a simple toy model and to the EGFR biochemical network from [1], the aim being to assess its accuracy. We solve in parallel the projected equations from [26], to verify explicitly the equivalence between the two approaches (under the condition specified above of negligible noise) to $\mathcal{O}(\delta x^2)$, and to compare the errors they make at higher order.

Everywhere in what follows we take $\epsilon \to 0$ (which implies also $\boldsymbol{\Sigma}^{\rm bb} = 0$) and the bulk is chosen initially at steady state, so $\boldsymbol{C}^{\rm bb|s}(t,t')$, given by (5.10), vanishes and $\boldsymbol{\mu}^{\rm b|s}(t) \equiv \boldsymbol{\nu}(t)$ (the conditional bulk means coincide with the deterministic bulk solution of the linearized dynamics). Furthermore, the coloured noise of the GVA is identically zero, as is the random force in the projection methods, at least up to $\mathcal{O}(\delta \boldsymbol{x}^2)$ (see Appendix E.2).

6.1. Toy model

To explain our method step by step we consider first a toy model with five species: x_1, x_2 and x_3 belong to the subnetwork, x_4 and x_5 are in the bulk, and they undergo two reversible complex formation/dissociation reactions. Schematically we write

$$x_1 + x_2 \underset{k^-_{3,12}}{\overset{k^+_{12,3}}{\rightleftarrows}} x_3$$

$$x_1 + x_4 \underset{k^-_{5,14}}{\overset{k^+_{14,5}}{\rightleftarrows}} x_5$$

where $k^+_{12,3}$ and $k^+_{14,5}$ are the complex formation rate constants, while $k^-_{3,12}$ and $k^-_{5,14}$ are the dissociation rates. The concentration deviations from the steady states $\delta x_i = x_i - \mu_i$ $i = 1, ..., 5$ follow the mass action kinetics

$$\begin{aligned}\frac{d}{dt}\delta x_1 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) \\ &\quad + k^-_{5,14}\delta x_5 - k^+_{14,5}(\mu_1\delta x_4 + \mu_4\delta x_1 + \delta x_1\delta x_4) \qquad (6.1a)\\ \frac{d}{dt}\delta x_2 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) \;=\; -\frac{d}{dt}\delta x_3 \qquad (6.1b)\\ \frac{d}{dt}\delta x_4 &= k^-_{5,14}\delta x_5 - k^+_{14,5}(\mu_1\delta x_4 + \mu_4\delta x_1 + \delta x_1\delta x_4) \;=\; -\frac{d}{dt}\delta x_5 \qquad (6.1c)\end{aligned}$$

There are no constant terms on the r.h.s. here as the μ_i describe a steady state. There are two conservation laws, one for the bulk and one for the subnetwork: the total concentration of x_2 and x_3 (similarly for x_4 and x_5) is constant in time, implying that $\delta x_2 = -\delta x_3$ and $\delta x_4 = -\delta x_5$. The subnetwork dynamics is therefore described only in terms of δx_1 and δx_2, and the bulk by δx_4, so that effectively $N^{\rm s} = 2$, $N^{\rm b} = 1$.

6.1.1. Reduced dynamics. There is one boundary species, δx_1, which interacts with the bulk species δx_4 and δx_5. Its dynamics is thus the only one affected by memory effects. By applying the formulas for the memory (4.6) and (5.12) we obtain the effective equation for $\delta x_1(t)$

$$\begin{aligned}\frac{d}{dt}\delta x_1 &= k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) + \int_0^t dt' M_{11}(t-t')\delta x_1(t') \\ &\quad + \int_0^t dt' \int_{t'}^t dt'' M_{11,1}(t,t',t'')\delta x_1(t')\delta x_1(t'') \qquad (6.2)\end{aligned}$$

with

$$M_{11}(t-t') = (k^-_{5,14} + k^+_{14,5}\mu_1)k^+_{14,5}\mu_4\, e^{-(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t')} \qquad (6.3)$$

and

$$M_{11,1}(t,t',t'') = k^{+2}_{14,5}\mu_4\, e^{-(k^-_{5,14}+k^+_{14,5}\mu_1)(t-t'')}[\delta(t-t') - (k^-_{5,14} + k^+_{14,5}\mu_1)] \qquad (6.4)$$

As explained in Sec. 5.2, solving the integro-differential equation (6.2) in differential form requires $2N^{\rm b} = 2$ additional variables. One of these just gives the mean of δx_4 conditional on δx_1

$$\frac{d}{dt}\delta\mu_{4|1} = -k^+_{14,5}\mu_4\delta x_1 - (k^-_{5,14} + k^+_{14,5}\mu_1)\delta\mu_{4|1}, \qquad \delta\mu_{4|1}(0) = 0 \tag{6.5}$$

The other one, z, using the procedure outlined in Appendix D, is the solution of

$$\frac{d}{dt}z = -k^+_{14,5}\delta x_1\delta\mu_{4|1} - (k^-_{5,14} + k^+_{14,5}\mu_1)z, \qquad z(0) = 0 \tag{6.6}$$

where $-(k^-_{5,14} + k^+_{14,5}\mu_1)$ is the only nonzero eigenvalue of $\boldsymbol{K}^{\rm bb}$. These additional equations must be solved jointly with the dynamics of the interior subnetwork species δx_2, which is given by the original (6.1b), and the version of (6.2) with memory terms expressed in terms of auxiliary variables

$$\begin{aligned}\frac{d}{dt}\delta x_1 = {} & k^-_{3,12}\delta x_3 - k^+_{12,3}(\mu_1\delta x_2 + \mu_2\delta x_1 + \delta x_1\delta x_2) \\ & - (k^-_{5,14} + k^+_{14,5}\mu_1)\delta\mu_{4|1} - k^+_{14,5}\delta x_1\delta\mu_{4|1} + (k^-_{5,14} + k^+_{14,5}\mu_1)z\end{aligned} \tag{6.7}$$

We solved (6.1b), (6.5), (6.6) and (6.7) to find the time evolution of $\delta x_1(t)$. Fig. 1 (left) shows the corresponding fractional concentration deviation from steady state, defined as $\delta\tilde{x}_1 = (x_1(t) - \mu_1)/\mu_1$, alongside the baseline obtained from the full reaction equations (black solid line): visually the two are essentially indistinguishable. Here and in what follows we refer to the "full reaction" equations as the system of Langevin equations describing the evolution of the *entire* network (comprising both subnetwork and bulk), taken in the limit of negligible noise ($\epsilon \to 0$) as for the reduced model. We contrast this with two simpler approximate subnetwork reductions for biochemical networks that are Markovian, i.e. do not involve memory terms: one in which the subnetwork is considered as isolated, i.e. reactions with the bulk are ignored, and one in which the bulk is assumed to be fast, thus always in its steady state given the instantaneous subnetwork concentrations (steady state bulk). Fig. 1 (left) demonstrates that both of these are distinctly less accurate than our nonlinear GVA reduction. The two simpler approximation schemes are typically justified either by timescale separation between subnetwork and bulk, or by weak subnetwork-bulk coupling. In our case the latter can be achieved by, for example, scaling down $k^+_{14,5}$ and $k^-_{5,14}$ by a common factor; as this factor is made large, we find indeed that all the approximate schemes perform well, making identical predictions in the limit. If on the other hand we increase the rates and hence the subnetwork-bulk coupling, the two alternative approximations, in particular the steady state bulk method, lead to increasingly poor predictions (see inset of Fig. 1, left). A systematic approach to model reduction, as provided by our method and also by projection approaches [26], gives good agreement with the solution of the full reaction equations *regardless* of any assumption on the strength of subnetwork-bulk interactions. This makes it more flexible in applications, leaving the choice of subnetwork free according to e.g. available data or biological interest in certain molecular species.

6.1.2. Quantitative tests. We next look at the error of the various subnetwork reduction methods. As in [26] we vary the scale of the initial deviations from steady state, which we characterize by the root mean squared deviation $\delta = (\{\sum_s[\delta\hat{x}_s(0)]^2\}/N^{\rm s})^{1/2}$. We quantify the accuracy of the approximation in terms of

$$\Delta = \frac{1}{T}\int_0^t dt' \frac{1}{N^{\rm s}}\sum_{s=1}^{N^{\rm s}} |\delta\tilde{x}_s(t) - \delta\hat{x}_s(t)| \tag{6.8}$$

where $\delta\hat{x}_s(t)$ is the exact solution. Δ is an absolute deviation in the dimensionless concentration of each subnetwork species, averaged over species and a time interval of $T = 15$ s, chosen here to cover the transient regime of relaxation to the steady state. From Fig. 1 (right) one sees that the errors for the simpler Markovian approximations are substantially larger in absolute terms, by several orders of magnitude for moderate δ; this emphasizes the importance of memory terms for accurate predictions. The errors also grow *linearly* with δ, while they scale *cubically* both for nonlinear projection methods and GVA. The conclusion, supported by the theoretical analysis in Appendix E, is that by retaining nonlinear memory terms both methods

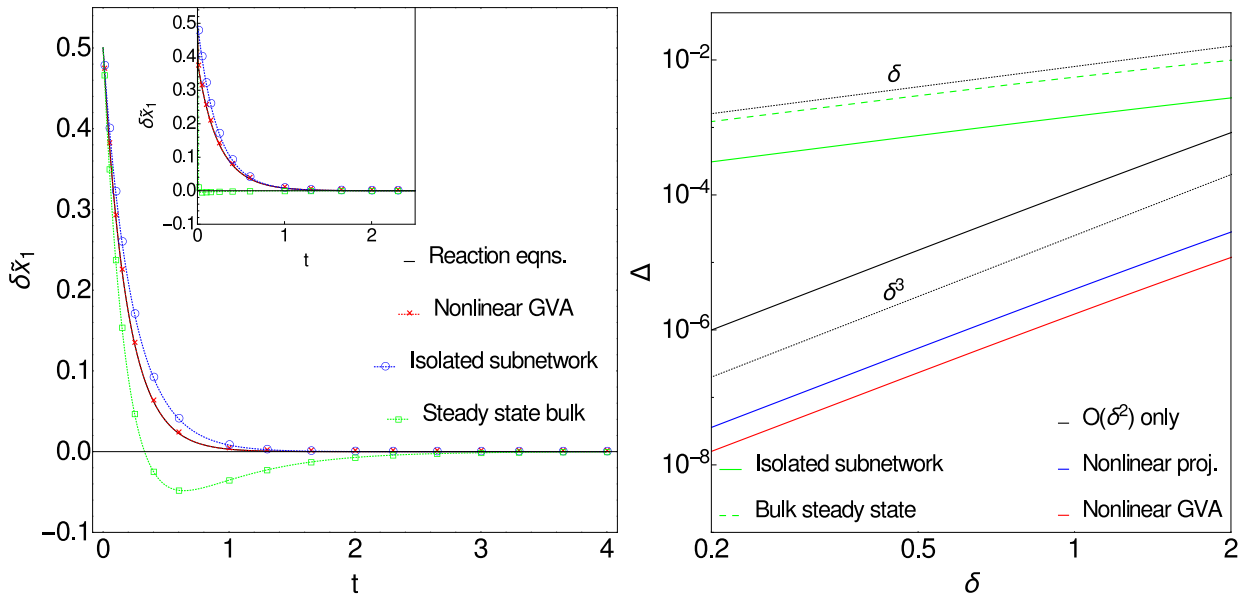

Figure 1. Toy model (Left) Time courses of $\delta\tilde{x}_1$ given by the nonlinear GVA, the isolated subnetwork and steady state bulk approximations compared to the full reaction equations. (Right) Approximation error, Δ, as a function of the initial deviation from steady state δ; the nonlinear GVA is compared to projection methods (both giving errors scaling as δ^3) and the two simpler approximations (whose error grows already as δ). Rates are set to $k^+_{12,3} = k^+_{14,5} = 1$, $k^-_{3,12} = k^-_{5,14} = 2$, giving steady state concentrations $\mu_1 = \mu_2 = \mu_4 = 1$, $\mu_3 = \mu_5 = 1/2$. In the inset on the left, the subnetwork-bulk coupling is increased by setting $k^+_{14,5} = 5$, $k^-_{5,14} = 10$. Initial conditions are chosen as $\delta\tilde{x}_1(0) = \delta\tilde{x}_2(0) = 1/2$, $\delta\tilde{x}_3(0) = -1$, $\delta\tilde{x}_4(0) = \delta\tilde{x}_5(0) = 0$ (bulk is initially at steady state) on the left figure. On the right we scale these initial conditions to get the chosen δ.

make predictions accurate to $\mathcal{O}(\delta x^2)$. Each method also includes a subset of higher order terms, but not systematically. At $\mathcal{O}(\delta x^3)$ the two methods therefore differ, and this leads to different prefactors in the scaling of their prediction error with δ^3.

In the nonlinear projection method, the error at $\mathcal{O}(\delta x^3)$ is due to the fact that there are cubic contributions to the random force that do not vanish even if the bulk is initially at steady state. In the nonlinear GVA, the error arises from the truncation of the perturbative expansion after $\mathcal{O}(\Delta\mathcal{H})$, while to be systematic in the dynamics at cubic order in δx one would have to retain also *fourth* order terms $\sim \hat{x}(\delta x)^3$ in the effective action, which would be affected by the $\mathcal{O}(\Delta\mathcal{H}^2)$ term in the expansion. Both methods include *some* higher order terms, and these do improve predictions: the curve labelled "$\mathcal{O}(\delta x^2)$ only" in Fig. 1 (right) shows that errors increase by more than an order of magnitude if one uses a reduction method that systematically discards all terms beyond quadratic order. (This is done in practice by evaluating all product terms in the reduced evolution equations from the linearized dynamics only.)

6.2. Application to EGFR

We consider the network of protein-protein interactions around EGFR as in the model by Kholodenko et al. [1]. This has previously been used as a testbed for projection methods in [26]; we refer to this work for a characterization of the reaction network, including parameters and the choice of initial conditions.

6.2.1. Quantitative tests. As before we take the limit of vanishing intrinsic noise, and initial conditions with the bulk at steady state. As shown in Fig. 2 (left), a very accurate prediction of subnetwork time courses is achieved using the nonlinear GVA reduction method. For a more quantitative assessment we again vary the initial deviation δ from steady state and measure the approximation error Δ defined by (6.8) (here we consider $T = 150$ s as time window for the transient regime as in [26]). The results are

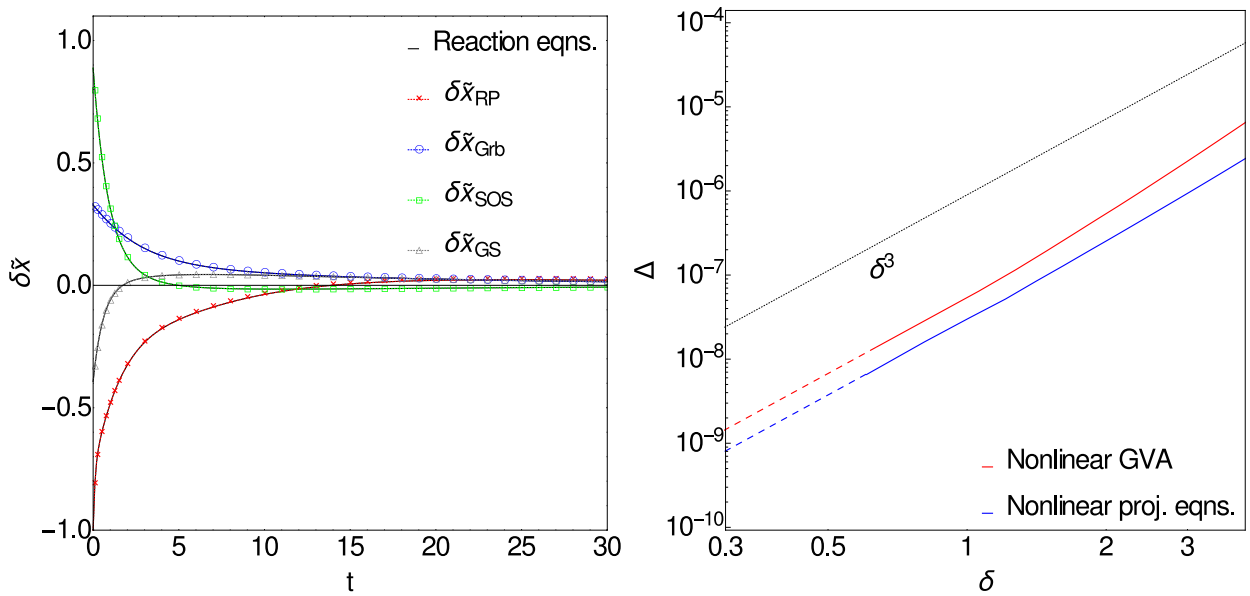

Figure 2. EGFR network (Left) Time courses of the fractional concentration deviations from steady state for the four subnetwork boundary species (defined as in [26]), as predicted by the nonlinear GVA reduction versus the full reaction equations. (Right) Approximation error, Δ, as a function of the initial deviation from steady state δ; the error is shown for both the nonlinear GVA and the nonlinear projected equations; for both the error scales as δ^3 as long as δ is not too large. The dashed lines are an extrapolation into the regime where errors become too small to measure reliably. Rates, steady states and initial conditions are set as in [26].

displayed in Fig. 2 (right), where we compare the approximation error of the the nonlinear GVA and the nonlinear projection methods. As expected, Δ grows as δ^3 for both methods. Again there is a difference in the prefactor, due to different terms being included beyond quadratic order. For the EGFR network, it is the projected subnetwork equations that are more accurate, by a factor of roughly two, whereas for the toy model it was the nonlinear GVA that produced smaller errors, cf. Fig. 1 (right). Which method is more accurate is therefore likely to be system-dependent in general. However, the difference in their errors is modest, while both methods are orders of magnitude more accurate than simpler Markovian approximations (see data in [26] for the EGFR case). When deciding which method for subnetwork reduction to use one might therefore prioritize the computational advantage offered by the nonlinear GVA over the (potentially, depending on the system) slightly higher accuracy of projection methods. As an example, for the EGFR case, a numerical implementation of the nonlinear GVA reduction method in terms of differential equations requires only 20 additional variables, as opposed to 255 for the nonlinear projected equations (see Appendix D and [26]).

6.2.2. Nonlinear memory function. We next analyze the nonlinear memory function $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'') \equiv \boldsymbol{M}^{\mathrm{s,ss}}(t-t',t-t'')$. We have defined this in (5.12) for ordered times $t' > t''$; for visual purposes it is helpful to plot the symmetrized version, which sets $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'') = \boldsymbol{M}^{\mathrm{s,ss}}(t,t'',t')$. The first term in (5.12) contains a factor $\delta(t-t')$ and thus gives contributions only along the "edge" $t' = t$ of the allowed region of times $(t'',t' < t)$. The remaining terms contribute in the "interior" of this region. We therefore denote these memory function contributions as GVA-edge and GVA-interior, respectively. For graphical purposes we plot the edge term simply as a function of $t-t''$, dropping the δ factor – see Figs. 3 and 4 (left) – while examples of the interior terms are shown in the 3D Figs. 3 and 4 (right). In particular, we consider a version of the nonlinear memory function in (5.12) made dimensionless w.r.t. concentrations, as it would appear in the dynamical subnetwork equations for the fractional concentration fluctuations $\delta\tilde{x}$; explicitly, $\tilde{M}_{i,jk}(t-t',t-t'') = y_i^{-1} M_{i,jk}(t-t',t-t'') y_j y_k$.

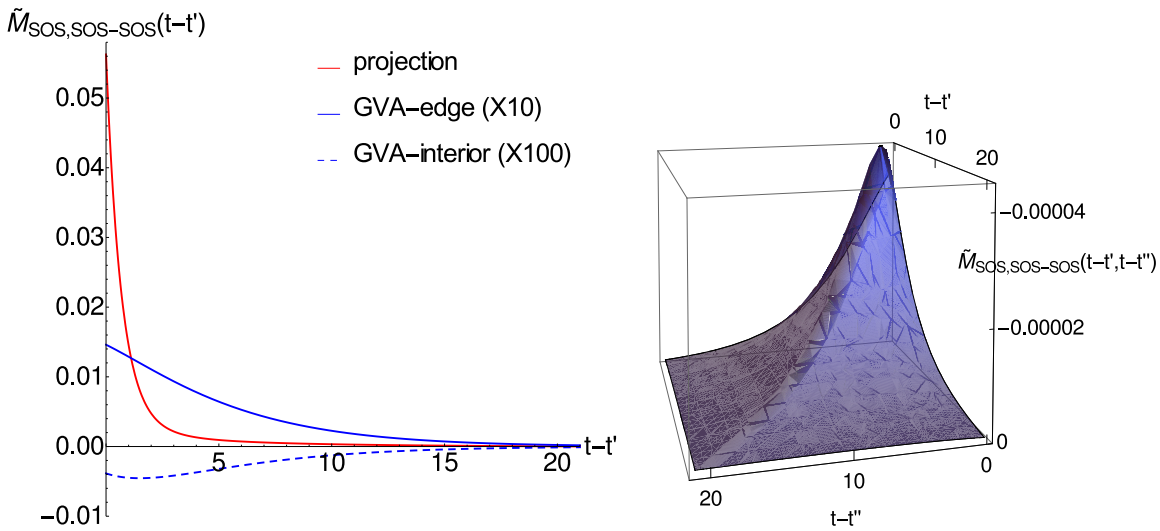

Figure 3. Nonlinear self-memory of subnetwork species SOS. (Left) Comparison of GVA-edge and GVA-interior memory function terms with nonlinear memory function from projection method; the GVA-interior term is shown along the diagonal, i.e. for $t' = t''$; the edge term is plotted against $t - t'$. The GVA-edge and interior terms are multiplied respectively by a factor of 10 and 100 to make them visible on the scale of the projection memory, whose amplitude is significantly bigger. Note though that only the projection memory and the GVA-edge memory contribution are comparable, having measurement units of $(\text{time})^{-2}$, while the GVA-interior contribution has units $(\text{time})^{-3}$. (Right) The GVA-interior memory function is plotted in 3D to show its shape for generic $t' \neq t''$: the diagonal contribution (i.e. for $t' = t''$) is dominant.

We compare the GVA nonlinear memory function with the one of the projection approach (see (E.8) in Appendix E.2), which is a function of just one time difference $t - t'$ as it acts only on concentration products at the *same* time t'. In Figs. 3 and 4 (left) we plot two examples from the EGFR network, the self-memory and a cross-memory function for SOS, against the GVA memory terms for the same species (where we set $t' = t''$ in the interior piece). In our setting (bulk initially at steady state and vanishing intrinsic noise), we know that the overall GVA nonlinear memory acting on a certain species is equivalent, to $\mathcal{O}(\delta x^2)$, to the one from the projection method. But the two methods decompose this nonlinear memory differently into sums of contributions from the possible products of concentrations. In particular, the projection memory features products involving *all* subnetwork concentrations, while the GVA memory involves only products of *boundary* species concentrations. This helps explain the significant difference in amplitude visible in Fig. 3 (left), where the GVA memory functions are so small that in order to make them visible on the same scale as the projection memory function we had to scale them up by 10 or 100, respectively. Inspection of projection functions describing memory to other products of concentrations shows that these have both positive and negative sign and so largely cancel. In the GVA memory there are fewer terms, namely those involving products of concentrations of boundary species, hence weaker cancellations: thus the individual memory terms can be smaller as we observe.

Finally, in Fig. 5, we show the entire (dimensionless) nonlinear memory on SOS, $\tilde{\mathcal{M}}_{\text{SOS}}(t)$, as a function of time along a specific time course, namely the one shown in Fig. 2 (left). We assess the relative importance of the GVA-edge and GVA-interior terms by plotting their separate contributions to the memory integral. This shows that either can be dominant, depending on time t.

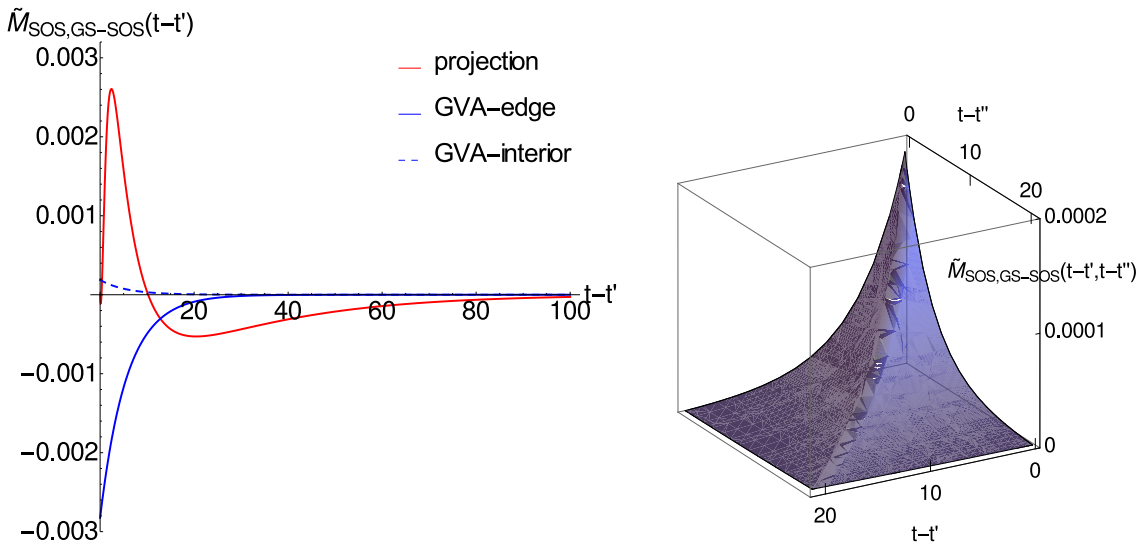

Figure 4. Example of nonlinear cross-memory acting on SOS, defined for the product GS-SOS. (Left) Comparison of GVA-edge and GVA-interior (for $t' = t''$) terms with projection methods memory. Similarly to $\tilde{M}_{\text{SOS,SOS-SOS}}(t-t')$ in Fig. 3 (left), the GVA nonlinear memory function $\tilde{M}_{\text{SOS,GS-SOS}}(t-t')$ is smaller in amplitude w.r.t. its projection analogue because effectively it accounts for what in the projection approach is a sum of terms with positive or negative sign - the nonlinear memories on SOS, defined for all subnetwork products but where one could imagine expressing the interior species in terms of GS and SOS (boundary ones). (Right) Full GVA-interior term as a function of $t-t'$ and $t-t''$: as in Fig. 3 (right), it is concentrated around $t' = t''$.

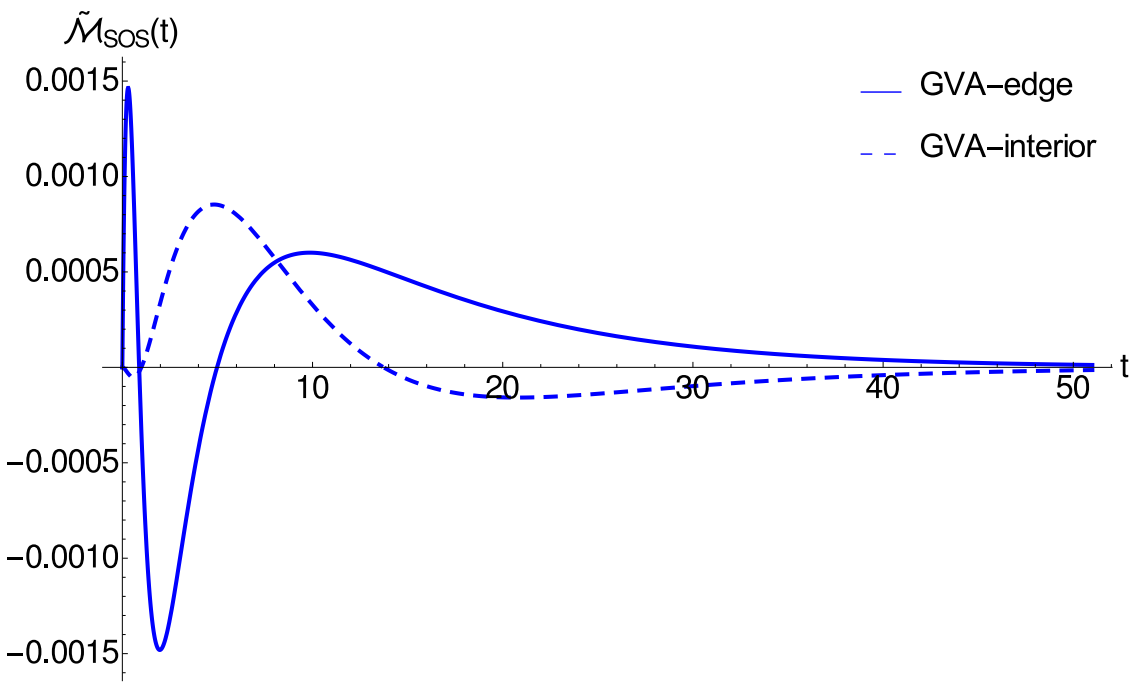

Figure 5. The integrals of the GVA-edge and GVA-interior memory terms, which together make up the entire nonlinear memory of SOS, as a function of time t for the time course from Fig. 2 (left). The two contributions are broadly comparable, and which one is dominant actually changes with time.

7. Discussion and conclusions

In this paper we have developed a model reduction technique based on a Gaussian approximation of the Chemical Langevin Equation (CLE) for biochemical networks with unary and binary reactions, describing e.g. complex formation and dissociation processes. We first re-derived a Gaussian Variational Approximation (GVA) via a path integral formalism. This is based on identifying the stationary point of a KL-divergence between complex distributions, and allows one to include the case of multiplicative noise in the CLE, see (3.10). One should be able to derive this result also using only real distributions as in [15]: the stationarity conditions for the KL should remain well-defined in the continuous time limit even if the KL itself formally diverges.

The GVA effectively yields a linearization of the dynamics around time-dependent means. The evolution of the latter (see (3.7)) includes corrections from the intrinsic noise that arises whenever the reaction volume V is finite. This gives the GVA a higher accuracy than the Linear Noise Approximation (LNA), which linearizes around the macroscopic, i.e. noise-free, time evolution. In particular the error of the GVA on the means can be shown to scale as $1/V^2$, which is already as small as the error one makes in any case by starting from the CLE rather than a master equation for discrete copy numbers. For the LNA, on the other hand, the error on the means is $\sim 1/V$ [22, 49].

From the GVA we derived reduced dynamical equations for a chosen subnetwork, by marginalizing out the other degrees of freedom. The coupling between the subnetwork and the rest of the network, or "bulk", leads to both a memory term and non-trivial time correlations in the effective noise. Evaluation of these quantities is feasible primarily when at least the bulk means are constant in time, which avoids the numerical calculation of time-ordered matrix exponentials. While the linearized approximation provided by the GVA may sound too simple, there are fully linear models e.g. in signalling pathways where they describe so-called *weakly activated* cascades and are considered of theoretical interest for building coarse-grained descriptions [50, 51].

We next developed a perturbative expansion of the effective action to estimate nonlinear corrections to the Gaussian approximation, and were able to derive from this explicit expression for the nonlinear memory functions. As a comparative baseline we used subnetwork equations derived by nonlinear projection techniques. There, nonlinear memory terms involve products of concentration fluctuations taken at the *same* time in the past, while in our new approach products at different time also appear. As we saw in two numerical examples in Sec. 6.2.2 the equal-time products nonetheless remain dominant.

We also compared our new nonlinear GVA subnetwork reduction scheme more quantitatively with the projection method. The latter can only be evaluated efficiently in the limit of small intrinsic noise, $\epsilon = 1/V \to 0$, so we focussed on this regime. We were able to show that the nonlinear GVA reaches the same level of accuracy as the projection approach, giving exact results up to $\mathcal{O}(\delta x^2)$ within an expansion in the overall deviation δx of the system from its steady state. Computationally, the new method is rather more efficient: the subnetwork equations can be integrated by mapping them to differential equations using a moderate (linear in the size of the bulk) number of auxiliary variables. The projection method, on the other hand, requires a number of auxiliary variables scaling quadratically with the number of bulk nodes.

We illustrated the reduction method first on a toy model that can be treated fully analytically. Numerical evaluation showed significantly more accurate predictions of time courses than can be achieved by simpler model reduction schemes (isolated subnetwork and steady state bulk). These work well primarily under appropriate assumptions of timescale separation, whereas our systematic derivation of a reduced model does not rely on any such assumption. The choice of subnetwork can therefore be adjusted flexibly, as required by whatever specific system one is studying.

Protein interaction networks are a key application for our model reduction strategy. We showed for such a model (the EGFR network from [1]) that the agreement with exact time courses is again excellent, the absolute error being slightly higher than for projection methods but still extremely small. We regard such improvements in prediction accuracy as important in constraining realistic models of biochemical networks and in helping to interpret experimental data correctly.

Interesting connections can be established to other approaches [31, 32] also based on *marginalizing* out the bulk or "environment". A practical limitation of such methods is that their implementation depends

on finding suitable approximate schemes for the marginal and conditional dynamics, which generally cannot be computed analytically. Here we instead use a Gaussian approximation for the conditional dynamics and evaluate non-Gaussian corrections to the marginal dynamics, allowing us to obtain closed form expressions for the effective non-Markovian terms (memory and coloured noise).

Model reduction techniques as developed here can be seen as a tool for partitioning a network consistently and systematically into separate modules that are easier to study. If memory functions can be represented by sufficiently simple approximations, e.g. exponentials, this can lead to significant *computational* savings in predicting the subnetwork dynamics, compared to simulating the entire network. At the same time there are *conceptual* benefits in studying how a cellular subsystem couples to the rest of the cellular environment. It has been recently shown, for example, that equilibrium-like domains can arise in non-equilibrium systems [52] in the form of generic bipartite graphs, a scenario akin to our subnetwork–bulk split.

In future work we plan to study the implications of our approach for noise modelling in biochemical systems and how these compare to e.g. projection methods. Here we can exploit the fact that in the covariance of the effective noise, see (B.7), we obtain an explicit decomposition, i.e. intrinsic noise, inherent to the random timing of biochemical events, and extrinsic noise due to the interaction with the environment [53–55].

One would also like to lift some of the restrictions in our approximations. E.g. for biochemical networks that exhibit bi- or multistable states or oscillations a Gaussian approximation with its single peak may be poor. Gaussian mixture models [56] could then be deployed instead, or one could use the general variational formalism of Eyink [8]. For systems where small copy number fluctuations are important, we are studying the application of our methods to path integral representations of the chemical master equation (e.g. [57–59], as discussed in [46]).

Also, there are types of nonlinearities in reaction networks that are not captured by unary and binary reactions, e.g. Michaelis-Menten (enzymatic) kinetics and Hill functions. To extend our formalism to such one could follow a procedure similar to the one used to treat Michaelis-Menten kinetics by projection methods [29]: introduce fast variables and binary reactions that in the limit of large rates reproduce the desired kinetics; apply the reduction method to this larger system; and finally take the large rate limit to get back to a reduced description for the original kinetics.

As is clear from this work, a path integral representation of biochemical dynamics has a structure that allows the application of variational principles and is convenient for formal manipulations such as the marginalization over the bulk; furthermore it offers a flexible framework for developing a perturbative theory for nonlinear problems. More generally, mathematical tools borrowed from field theories in physics, such as path integrals [30, 60] and Feynman diagrams [61], can yield more accurate prediction methods and an understanding of dynamical behaviours grounded in detailed descriptions of fundamental components. These results are promising and could motivate a systematic transfer of approaches from theoretical physics to systems biology that is still to be fully explored.

Finally, as a natural continuation of our analysis, one could address the inverse problem: given observed time courses for the subnetwork, can we use the insight into reduced subnetwork dynamics that we have established to infer properties of the bulk? If the structure of the bulk is known, one could aim to infer the bulk dynamics from the observed subnetwork behaviour. In our approach this is exactly what is being done by conditioning on the subnetwork degrees of freedom, with bulk dynamics estimated by Gaussian conditioning; an interesting challenge would be to incorporate nonlinear corrections. Other approximate approaches to this problem, based on variational methods [13–15, 62] or system size expansions [63], have turned out to be clearer conceptually than MCMC sampling inference methods for intracellular stochastic kinetics.

Acknowledgements

This work was supported by the Marie Curie Training Network NETADIS (FP7, grant 290038). We thank Reimer Kühn for insightful comments on possible future directions, Luca Dall'Asta and Heinz Köppl for suggesting several related papers. We are indebted to Katy Rubin and Edgar Herrera Delgado for making available their code which helped us with the numerical implementation of projection methods. PS acknowledges the stimulating research environment provided by the EPSRC Centre for Doctoral Training

in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, EP/L015854/1). BB acknowledges the Simons Foundation Grant No. 454953.

[1] Kholodenko B N, Demin O V, Moehren G and Hoek J B 1999 *J. Biol. Chem.* **274** 30169–30181
[2] Plotnikov A, Zehorai E, Procaccia S and Seger R 2011 *Biochimica et Biophysica Acta* **1813** 1619–1633
[3] Van Kampen N G 2007 *Stochastic Processes in Physics and Chemistry* (Elsevier, 3rd Edition)
[4] Radulescu O, Gorban A, Zinovyev A and Noel V 2012 *Front. Genet.* **3**
[5] Ackermann J, Einloft J, Nothen J and Koch I 2012 *J. Theor. Biol.* **315** 71–80
[6] Okino M and Mavrovouniotis M 1998 *Chem. Rev.* **98** 391–408
[7] Apri M, De Gee M and Molenaar J 2012 *J. Theor. Biol.* **304** 16–26
[8] Eyink G 1996 *Phys. Rev. E* **54**
[9] Sasai M and Wolynes P 2003 *PNAS* **100** 2374–2379
[10] Ohkubo J 2007 *JSTAT* **P09017**
[11] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Springer)
[12] Kullback S and Leibler R A 1951 *Ann. Math. Stat.* **22** 79–86
[13] Opper M and Sanguinetti G 2007 *Proceedings of NIPS*
[14] Cohn I, El-Hay T, Friedman N and Kupferman R 2009 *Proceedings of NIPS*
[15] Archambeau C, Cornford D, Opper M and Shawe-Taylor J 2007 *JMLR: Workshop and Conference Proceedings* **1** 1–16
[16] Bachschmid-Romano L, Battistin C, Opper M and Roudi Y 2016 *J. Phys. A: Math. Theor.* **49** 434003
[17] Whittle P 1957 *J.R. Stat. Soc.* **19** 268–281
[18] Bosia C, Pagnani A and Zecchina R 2013 *Plos One* **8**
[19] Lakatos E, Ale A, Kirk P D W and Stumpf M P H 2015 *J. Chem. Phys.* **143**
[20] Schnoerr D, Sanguinetti G and Grima R 2014 *J. Chem. Phys.* **141**
[21] Schnoerr D, Sanguinetti G and Grima R 2015 *J. Chem. Phys.* **143**
[22] Grima R 2012 *J. Chem. Phys.* **136**
[23] Chorin A J, Hald O H and Kupferman R 2000 *PNAS* **97** 2968–2973
[24] Chorin A J, Hald O H and Kupferman R 2006 *J. Scient. Comput.* **28**
[25] Thomas P, Grima R and Straube A 2012 *Phys. Rev. E* **86** 041110
[26] Rubin K J, Lawler K, Sollich P and Ng T 2014 *J. Theor. Biol.* **357** 245–267
[27] Zwanzig R 1961 *Phys. Rev.* **124**
[28] Rubin K 2014 *Dynamics of protein interaction subnetworks* Ph.D. thesis King's College London
[29] Rubin K J and Sollich P 2016 *J. Chem. Phys.* **144**
[30] Sinitsyn N, Hengartner N, Nemenman I and Press W 2009 *PNAS* **106** 10546–10551
[31] Zechner C, Unger M, Pelet S, Peter M and Koeppl H 2014 *Nature Methods* **11** 197–202
[32] Zechner C and Koeppl H 2014 *PLoS Comput. Biol.* **10** e1003942
[33] Gardiner C W 1985 *Handbook for Stochastic Methods* (Springer, 2nd edition)
[34] Van Kampen N G 1982 *J. Stat. Phys.* **24** 175–187
[35] Elf J and Ehrenberg M 2003 *Genome Res.* **13** 2475–2484
[36] Coolen A and Rabello S 2009 *JPCS* **197**
[37] Martin P, Siggia E and Rose H 1973 *Phys. Rev. A* **8** 423–436
[38] Janssen H K 1976 *Z. Phys. B: Cond. Mat.* **23** 377–380
[39] De Dominicis C 1978 *Phys. Rev. B* **18** 4913–4919
[40] Kamenev A 2011 *Field theory of non-equilibrium systems* (Cambridge University Press)
[41] Hertz J A, Roudi Y and Sollich P 2017 *J. Phys. A: Math. Theor* **50** 033001
[42] Ritort F and Sollich P 2003 *Advances in Physics* **52** 219 — 342
[43] Chorin A J, Hald O H and Kupferman R 2002 *Physica D* **166** 239–257
[44] Coolen A C C 2001 *Handbook of Biological Physics* vol 4 (Elsevier-Academic Press, Amsterdam) chap Statistical Mechanics of Recurrent Neural Networks II Dynamics
[45] Fischer K H and Hertz J A 1991 *Spin Glasses* (Cambridge Studies in Magnetism)
[46] Bravi B 2016 *Path integral approaches to subnetwork dynamics and inference* Ph.D. thesis King's College London
[47] Thomas P, Straube A and Grima R 2012 *BMC Syst. Biol.* **6**
[48] Bravi B and Sollich P 2016 *Arxiv preprint 1603.05538*
[49] Grima R, Thomas P and Straube A 2011 *J. Chem. Phys.* **135**
[50] Diaz M, Desikan R and Barahona M 2015 *Arxiv preprint 1112.0270v2*
[51] Herath N, Hamadeh A and Del Vecchio D 2015 *Proceedings of ACC*
[52] Saad D and Mozeika A 2013 *Phys. Rev. E* **87** 032131
[53] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 *Science* **297** 1183–1186
[54] Swain P, Elowitz M B and Siggia E D 2002 *PNAS* **99** 12795–12800
[55] Shahrezaei V, Olivier J F and Swain P 2008 *Mol. Syst. Biol.* **4**
[56] Chang H, Hemberg M, Barahona M, Ingber D and Huang S 2008 *Nature* **453** 544–547
[57] Doi M 1976 *J. Phys. A: Math. Gen.* **9** 1465–1477
[58] Doi M 1976 *J. Phys. A: Math. Gen.* **9** 1479–1495
[59] Peliti L 1985 *Journal de Physique* **46** 1469–1483
[60] Santos F A N, Gadelha H and Gaffney E A 2015 *Phys. Rev. E* **92** 062714
[61] Thomas P, Fleck C, Grima R and Popovic N 2014 *J. Phys. A: Math. Theor.* **47**
[62] Opper M and Archambeau C 2009 *Neural Comput.* **21** 786–92
[63] Ruttor A and Opper M 2009 *Phys. Rev. Lett.* **103** 230601
[64] Kleinert H 2006 *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* 4th ed (World scientific publishing)

Supplementary material

Appendix A. Gaussian Variational Approximation

We derive in this appendix the stationarity conditions for the KL divergence that define the Gaussian Variational Approximation (GVA), leading in particular to (3.7)-(3.10) in the main text.

We denote $\boldsymbol{y} = (\boldsymbol{x}, -\mathrm{i}\hat{\boldsymbol{x}})$ in such a way that the joint "distribution" of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ can be written

$$P(\boldsymbol{y}) = \frac{e^{\mathcal{H}(\boldsymbol{y})}}{Z} P_0(\boldsymbol{x}(0)) \tag{A.1}$$

$P_0(\boldsymbol{x}(0))$ being the unknown initial probability distribution. We then proceed with the variational approximation by assuming that the approximating distribution is Gaussian

$$P(\boldsymbol{y}) = \frac{e^{\mathcal{H}(\boldsymbol{y})}}{Z} P_0(\boldsymbol{x}(0)) \approx \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_{\mathrm{gen}}, \boldsymbol{C}_{\mathrm{gen}}) = Q(\boldsymbol{y}) \tag{A.2}$$

The Gaussian $\mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_{\mathrm{gen}}, \boldsymbol{C}_{\mathrm{gen}})$ is completely determined by two sets of parameters, the vector of mean values $\boldsymbol{\mu}_{\mathrm{gen}} = \langle \boldsymbol{y} \rangle_Q$ and the covariance matrix $\boldsymbol{C}_{\mathrm{gen}} = \langle \boldsymbol{y}\boldsymbol{y}^T \rangle_Q - \boldsymbol{\mu}_{\mathrm{gen}}\boldsymbol{\mu}_{\mathrm{gen}}^T$. We define the Kullback-Leibler (KL) Divergence [12] between Q and P

$$\begin{aligned}
\mathrm{KL}(Q||P) &= \int D\boldsymbol{y}\, Q(\boldsymbol{y}) \ln \frac{Q(\boldsymbol{y})}{P(\boldsymbol{y})} = \\
&= \left\langle -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{gen}})^T \boldsymbol{C}_{\mathrm{gen}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\mathrm{gen}}) - \frac{1}{2} \ln (2\pi)^d \det(\boldsymbol{C}_{\mathrm{gen}}) + \mathcal{H}(\boldsymbol{y}) + \ln Z - \ln P_0(\boldsymbol{x}(0)) \right\rangle_Q = \\
&= -\frac{d}{2} - \frac{1}{2} \ln \left[(2\pi)^d \det(\boldsymbol{C}_{\mathrm{gen}}) \right] + \ln Z - \left\langle \mathcal{H}(\boldsymbol{y}) + \ln P_0(\boldsymbol{x}(0)) \right\rangle_Q
\end{aligned} \tag{A.3}$$

where d is the dimension of the vectors involved, i.e. $d = N(2T/\Delta + 1)$ and Δ is the elementary time step. With real measures, $\mathrm{KL}(Q||P) \geq 0$ with equality if and only if $P(\boldsymbol{y}) \equiv Q(\boldsymbol{y})$, so it can be seen as a measure of the dissimilarity (a "distance") of the distributions $Q(\boldsymbol{y})$ and $P(\boldsymbol{y})$. This leads one to choose the variational parameters by minimizing the KL divergence, or more generally in the case of complex measures to make it stationary

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{\mu}_{\mathrm{gen}}} = 0 \tag{A.4a}$$

$$\frac{\partial \mathrm{KL}}{\partial \boldsymbol{C}_{\mathrm{gen}}} = 0 \tag{A.4b}$$

Given the KL expression (A.3), the set of equations (A.4a) for the components of $\boldsymbol{\mu}_{\mathrm{gen}}$ reduces to

$$\nabla_{\boldsymbol{\mu}_{\mathrm{gen}}} \left(\left\langle \mathcal{H}(\boldsymbol{y}) + \ln P_0(\boldsymbol{x}(0)) \right\rangle_Q \right) = 0 \tag{A.5}$$

while the equation for the inverse correlation matrix (A.4b) is

$$\frac{1}{2} \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \ln (\det \boldsymbol{C}_{\mathrm{gen}}) + \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \langle \mathcal{H}(\boldsymbol{y}) \rangle_Q + \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \langle \ln P_0(\boldsymbol{x}(0)) \rangle_Q = 0 \tag{A.6}$$

For the first term in (A.6), one can exploit the properties

$$\nabla_{\boldsymbol{C}_{\mathrm{gen}}} \ln (\det \boldsymbol{C}_{\mathrm{gen}}) = \nabla_{\boldsymbol{C}_{\mathrm{gen}}} \mathrm{Tr}(\ln \boldsymbol{C}_{\mathrm{gen}}) \tag{A.7}$$

$$\frac{\partial \,\mathrm{Tr}(\ln \boldsymbol{C}_{\mathrm{gen}})}{\partial (\boldsymbol{C}_{\mathrm{gen}})_{ijtt'}} = (\boldsymbol{C}_{\mathrm{gen}}^{-1})_{jit't} \tag{A.8}$$

and the following identities (see [62] for a proof)

$$\nabla_{\boldsymbol{\mu}_{\rm gen}} \langle \mathcal{H}(\boldsymbol{y})\rangle_Q = \langle \nabla_{\boldsymbol{y}} \mathcal{H}(\boldsymbol{y})\rangle_Q \tag{A.9a}$$

$$\nabla_{\boldsymbol{C}_{\rm gen}} \langle \mathcal{H}(\boldsymbol{y})\rangle_Q = \frac{1}{2} \langle \nabla_{\boldsymbol{y}} \nabla_{\boldsymbol{y}} \mathcal{H}(\boldsymbol{y})\rangle_Q \tag{A.9b}$$

To apply these, one has to calculate the derivatives of $\mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})$ with respect to the variables $\boldsymbol{x}$ and $-\mathrm{i}\hat{\boldsymbol{x}}$

$$\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = x_i(t+\Delta) - x_i(t) - \Delta\Phi_i(\boldsymbol{x}(t)) + \Delta \sum_j \mathrm{i}\hat{x}_j(t)\Sigma_{ji}(\boldsymbol{x}(t)) \tag{A.10}$$

$$\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = \mathrm{i}\hat{x}_i(t-\Delta) - \mathrm{i}\hat{x}_i(t) - \Delta \mathrm{i}\hat{x}_i(t)\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \frac{\Delta}{2}\sum_{jk} \mathrm{i}\hat{x}_j(t)\frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t) \tag{A.11}$$

with

$$\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} = \sum_{j,l,j\neq l} k^+_{ij,l} x_j(t) + \frac{1}{2}\sum_{j,l,j\neq l} k^-_{i,jl} + \sum_j \lambda_{ij} + \sum_l 2k^+_{ii,l} x_i(t) + \sum_j k^-_{i,jj} \tag{A.12}$$

For the second derivatives we need a number of different combinations

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = -\Delta \mathrm{i}\hat{x}_i(t)\frac{\partial}{\partial x_j(t)}\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \frac{\Delta}{2}\sum_{lk} \mathrm{i}\hat{x}_l(t)\frac{\partial}{\partial x_j(t)}\frac{\partial \Sigma_{lk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t) \tag{A.13a}$$

$$\frac{\partial}{\partial(\mathrm{i}\hat{x}_i(t))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = -1 - \Delta\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)} + \Delta\sum_k \frac{\partial \Sigma_{ik}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t) \tag{A.13b}$$

$$\frac{\partial}{\partial(\mathrm{i}\hat{x}_j(t))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = \Delta\Sigma_{ij}(\boldsymbol{x}(t)) \tag{A.13c}$$

$$\frac{\partial}{\partial x_j(t)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = -\Delta\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)} + \Delta\sum_k \mathrm{i}\hat{x}_k(t)\frac{\partial \Sigma_{ki}(\boldsymbol{x}(t))}{\partial x_j(t)} \qquad i\neq j \tag{A.13d}$$

$$\frac{\partial}{\partial x_i(t+\Delta)}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} = \frac{\partial}{\partial(\mathrm{i}\hat{x}_i(t-\Delta))}\frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} = 1 \tag{A.13e}$$

For all other time differences the second derivatives are zero.

We can now obtain the explicit equations for the means. Considering the $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ variables separately, one can cast the equations (A.4a) in the form

$$\left\langle \frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial x_i(t)} + \frac{\partial(\ln P_0(\boldsymbol{x}(0))}{\partial x_i(t)}\right\rangle = 0 \qquad \left\langle \frac{\partial \mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})}{\partial(\mathrm{i}\hat{x}_i(t))} + \frac{\partial(\ln P_0(\boldsymbol{x}(0)))}{\partial(\mathrm{i}\hat{x}_i(t))}\right\rangle = 0 \tag{A.14}$$

The initial distribution $P_0(\boldsymbol{x}(0))$ only contributes in the left equation, for $t=0$. We make the natural choice of assuming it to be Gaussian, $P_0(\boldsymbol{x}(0)) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_0,\boldsymbol{C}_0)$ so that

$$\left\langle \frac{\partial \ln P_0}{\partial x_i(0)}\right\rangle = -\sum_j (\boldsymbol{C}_0^{-1}(0))_{ij}(\boldsymbol{\mu}(0)-\boldsymbol{\mu}_0(0))_j \tag{A.15}$$

Then given the derivatives (A.10) and (A.11) and carrying out the average in (A.14) one obtains

$$\mathrm{i}\hat{\mu}_i(t) - \mathrm{i}\hat{\mu}_i(t-\Delta) + \Delta\left\langle \mathrm{i}\hat{x}_i(t)\frac{\partial \Phi_i(\boldsymbol{x}(t))}{\partial x_i(t)}\right\rangle - \frac{\Delta}{2}\left\langle \sum_{jk} \mathrm{i}\hat{x}_j(t)\frac{\partial \Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t)\right\rangle = 0 \tag{A.16}$$

where the term (A.15) has to be added on the l.h.s. for $t=0$. For the ordinary means, on the other hand, one has

$$\mu_i(t+\Delta) - \mu_i(t) - \Delta\langle \Phi_i(\boldsymbol{x}(t))\rangle + \Delta\left\langle \sum_j \mathrm{i}\hat{x}_j(t)\Sigma_{ji}(\boldsymbol{x}(t))\right\rangle = 0 \tag{A.17}$$

with

$$\begin{aligned}\langle\Phi_i(\boldsymbol{x}(t))\rangle = \Phi_i(\boldsymbol{\mu}(t),\boldsymbol{C}(t,t)) = \sum_{j,l,j\neq l}\Big(k^-_{l,ij}\mu_l(t) - k^+_{ij,l}(\mu_i(t)\mu_j(t)+C_{ij}(t,t))\Big) + \\ + \frac{1}{2}\sum_{j,l,j\neq l}\Big(k^+_{jl,i}(\mu_j(t)\mu_l(t)+C_{jl}(t,t)) - k^-_{i,jl}\mu_i(t)\Big) + \sum_j\big(\lambda_{ji}\mu_j(t)-\lambda_{ij}\mu_i(t)\big) + \\ + \sum_l\Big(2k^-_{l,ii}\mu_l(t) - k^+_{ii,l}(\mu_i(t)\mu_i(t)+C_{ii}(t,t))\Big) + \sum_j\left(\frac{1}{2}k^+_{jj,i}(C_{jj}(t,t)+\mu_j(t)\mu_j(t)) - k^-_{i,jj}\mu_i(t)\right)\end{aligned} \tag{A.18}$$

The "initial" (boundary) condition for (A.16) at $t=T$, given by $\langle\partial\mathcal{H}/\partial x_i(T)\rangle = 0$, yields $\hat{\mu}_i(T-\Delta)=0$. If we also assume that the approximate solution is causal (i.e. $R_{ij}(t,t) = -\mathrm{i}\langle\hat{x}_i(t)x_j(t)\rangle = 0$) and that $-\langle\hat{x}_i(t)\hat{x}_j(t)\rangle \equiv 0$, then the two averages in (A.16) vanish term by term in a Taylor expansion in $\boldsymbol{x}(t)$, and solving backwards in time we have that $\hat{\mu}_i(t)=0\ \forall t$. At the initial time this then requires $\boldsymbol{\mu}(0)=\boldsymbol{\mu}_0(0)$ to make also the term (A.15) zero, which is intuitively reasonable.

Note that in the exact trajectory distribution, the above assumptions on averages clearly hold: averages of any product of the $\hat{x}_i(t)$ vanish, as a consequence of the normalization of the overall trajectory distribution (as shown in [45]), while the vanishing of the response function $-\mathrm{i}\langle\hat{x}_i(t)x_j(t')\rangle$ for $t>t'$ follows from causality and, at $t=t'$, from the fact that we are using an Itô discretization [41]. It then makes sense that a Gaussian approximation to the trajectory distribution would maintain these properties.

In equation (A.17), vanishing of the equal-time response again makes the last term zero. Rearranging and taking the limit $\Delta\to 0$ one obtains

$$\frac{d\mu_i(t)}{dt} = \Phi_i(\boldsymbol{\mu}(t),\boldsymbol{C}(t,t)) \tag{A.19}$$

This differs from the deterministic equation of motion for the mean values by additional terms stemming from equal time (co-)variances $\boldsymbol{C}(t,t)$.

Moving on to the variational equations for the (generalized) covariance matrix, one starts from (A.6) and uses the property (A.9b) as well as the explicit second derivatives (A.13) of $\mathcal{H}(\boldsymbol{x},\hat{\boldsymbol{x}})$. This shows that nonzero inverse correlation elements, for $t\neq 0$, are given by

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{jitt} = \Delta\mathrm{i}\hat{\mu}_i(t)\sum_l k^+_{ij,l} - \frac{\Delta}{2}\left\langle\sum_{lk}\mathrm{i}\hat{x}_l(t)\frac{\partial}{\partial x_j(t)}\frac{\partial\Sigma_{lk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t)\right\rangle \qquad i\neq j \tag{A.20a}$$

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{iitt} = \Delta\mathrm{i}\hat{\mu}_i(t)\sum_l 2k^+_{ii,l} - \frac{\Delta}{2}\left\langle\sum_{jk}\mathrm{i}\hat{x}_j(t)\frac{\partial}{\partial x_i(t)}\frac{\partial\Sigma_{jk}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t)\right\rangle \tag{A.20b}$$

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{\hat{\imath}itt} = -1-\Delta K_{ii}(t) + \left\langle\Delta\sum_k\frac{\partial\Sigma_{ik}(\boldsymbol{x}(t))}{\partial x_i(t)}\mathrm{i}\hat{x}_k(t)\right\rangle \tag{A.20c}$$

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{\hat{\imath}\hat{\jmath}tt} = -\Delta\langle\Sigma_{ij}(\boldsymbol{x}(t))\rangle \tag{A.20d}$$

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{j\hat{\imath}tt} = -\Delta K_{ij}(t) + \left\langle\Delta\sum_k\mathrm{i}\hat{x}_k(t)\frac{\partial\Sigma_{ki}(\boldsymbol{x}(t))}{\partial x_j(t)}\right\rangle \qquad i\neq j \tag{A.20e}$$

$$(\boldsymbol{C}^{-1}_{\mathrm{gen}})_{i\hat{\imath}t+\Delta t} = (\boldsymbol{C}^{-1}_{\mathrm{gen}})_{\hat{\imath}it-\Delta t} = 1 \tag{A.20f}$$

where the indexes $\hat{\imath}$ and $\hat{\jmath}$ are introduced to refer to $\hat{x}_i$ and $\hat{x}_j$. For $t=0$, there would be an extra term on the r.h.s. of (A.20a) from

$$\left\langle\frac{\partial}{\partial x_j(0)}\frac{\partial\ln P_0(\boldsymbol{x}(0))}{\partial x_i(0)}\right\rangle = -(\boldsymbol{C}^{-1}(0))_{ij} \tag{A.21}$$

which just sets the initial covariance matrix to be the exact one. We have defined the quantity $K_{ij}(t)$ as follows

$$K_{ij}(t) = \left\langle\frac{\partial\Phi_i(\boldsymbol{x}(t))}{\partial x_j(t)}\right\rangle \tag{A.22}$$

From the expressions for the inverse correlation matrix we now want to recover differential equations for the temporal evolution of the correlations. This can be done by imposing the inverse relation

$$\sum_{jt''} \left[(\boldsymbol{C}_{\text{gen}})_{kjtt''} (\boldsymbol{C}_{\text{gen}})^{-1}_{jit''t'} + (\boldsymbol{C}_{\text{gen}})_{k\hat{j}tt''} (\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{j}it''t'} \right] = \delta_{ki}\delta_{tt'} \tag{A.23}$$

If we consider explicitly only the nonzero terms in the summation over t'', the sum becomes

$$\begin{aligned} &\sum_{j\neq i} (\boldsymbol{C}_{\text{gen}})_{kjtt'} (\boldsymbol{C}_{\text{gen}})^{-1}_{jit't'} + (\boldsymbol{C}_{\text{gen}})_{kitt'} (\boldsymbol{C}_{\text{gen}})^{-1}_{iit't} + \sum_{j\neq i} (\boldsymbol{C}_{\text{gen}})_{k\hat{j}tt'} (\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{j}it't} + \\ &+ (\boldsymbol{C}_{\text{gen}})_{k\hat{i}tt'} (\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{i}it't} + (\boldsymbol{C}_{\text{gen}})_{k\hat{i}tt'-\Delta} (\boldsymbol{C}_{\text{gen}})^{-1}_{\hat{i}it'-\Delta t'} = \delta_{kitt'} \end{aligned} \tag{A.24}$$

As before, we can use the fact that we the solution will be causal and all averages of products of auxiliary variables to vanish. Substituting the expressions (A.20) into (A.23), dividing by Δ and taking the limit for $\Delta \to 0$ then yields the differential equations

$$\frac{\partial R_{ki}(t,t')}{\partial t'} = -\delta_{ki}\delta(t-t') - \sum_j R_{kj}(t,t') K_{ji}(t') \tag{A.25}$$

$$\frac{\partial R_{ki}(t,t')}{\partial t} = \delta_{ki}\delta(t'-t) + \sum_{j\neq k} K_{kj}(t) R_{ji}(t,t') \tag{A.26}$$

$$\frac{\partial C_{ki}(t,t')}{\partial t'} = \sum_j C_{kj}(t,t') K_{ij}(t') + \sum_j R_{kj}(t,t') \langle \Sigma_{ji}(\boldsymbol{x}(t')) \rangle \tag{A.27}$$

As discussed in the main text, the results of the GVA can be understood as an effective linearization of the Langevin dynamics (2.1) around time dependent means. Let us show in more detail that the equations (A.25), (A.26), (A.27) can be mapped into this case. We shall start from the fluctuations about the mean, which obey

$$\frac{d(\boldsymbol{x}(t) - \boldsymbol{\mu}(t))}{dt} = \boldsymbol{K}(t)(\boldsymbol{x}(t) - \boldsymbol{\mu}(t)) + \boldsymbol{\xi}(t) \tag{A.28}$$

$\boldsymbol{K}(t)$ represents a time dependent rate matrix (as given by the linearization around time-dependent means (A.22)) and the noise satisfies $\langle \boldsymbol{\xi}(t)\boldsymbol{\xi}^T(t') \rangle = \langle \boldsymbol{\Sigma}(\boldsymbol{x}(t)) \rangle \delta(t-t')$. The general solution reads

$$\boldsymbol{x}(t) = \boldsymbol{\mu}(t) + \boldsymbol{R}(t,0)(\boldsymbol{x}(0) - \boldsymbol{\mu}(0)) + \int_0^t ds\, \boldsymbol{R}(t,s)\boldsymbol{\xi}(s) \tag{A.29}$$

in terms of the response function

$$\boldsymbol{R}(t,t') = \theta(t-t')\, \mathrm{T} \left[e^{\int_{t'}^{t} \boldsymbol{K}(t'')dt''} \right] \tag{A.30}$$

Here the time-ordering operator T indicates that in the expansion of the exponential the matrices $\boldsymbol{K}(t'')$ are always ordered in order of ascending time from right to left. Applying the definition of correlations one then also has

$$\boldsymbol{C}(t,t') = \boldsymbol{R}(t,0)\boldsymbol{C}(0,0)\boldsymbol{R}^T(t',0) + \int_0^{\min(t,t')} ds\, \boldsymbol{R}(t,s) \langle \boldsymbol{\Sigma}(\boldsymbol{x}(s)) \rangle \boldsymbol{R}^T(s,t') \tag{A.31}$$

It is straightforward to check that these response and correlation functions satisfy the equations (A.25), (A.26), (A.27) that we had obtained variationally.

Appendix B. Memory and coloured noise in the linearized dynamics

Appendix B.1. Non-stationary case

We start from the Langevin equations for concentrations deviations from the means $\delta\boldsymbol{x}^{\rm s}(t)$ and $\delta\boldsymbol{x}^{\rm b}(t)$ given by (4.1) and (4.2) of the main text, obtained via a linearization of the dynamics around possibly time-dependent means $\boldsymbol{\mu}^{\rm s}(t)$ and $\boldsymbol{\mu}^{\rm b}(t)$. For simplicity, we set the subnetwork-bulk blocks of the initial conditions and of fluctuations to zero, i.e. $\boldsymbol{C}^{\rm sb}(0,0) = \langle\delta\boldsymbol{x}^{\rm s}(0)\delta\boldsymbol{x}^{\rm b\,T}(0)\rangle = \mathbb{0}$ (and similarly for $\boldsymbol{C}^{\rm bs}(0,0)$), $\langle\boldsymbol{\Sigma}^{\rm sb}(\boldsymbol{x}(t))\rangle = \langle\boldsymbol{\xi}^{\rm s}(t)\boldsymbol{\xi}^{\rm b\,T}(t)\rangle = \mathbb{0}$ (and similarly for $\langle\boldsymbol{\Sigma}^{\rm bs}(\boldsymbol{x}(t))\rangle$): this assumption can be easily lifted at the price of longer expressions.

We want to read off the structure of the reduced dynamics for $\delta\boldsymbol{x}^{\rm s}(t)$, which can be cast as

$$\frac{d\delta\boldsymbol{x}^{\rm s}(t)}{dt} = \boldsymbol{K}^{\rm ss}(t)\delta\boldsymbol{x}^{\rm s}(t) + \int_0^t \boldsymbol{M}^{\rm ss}(t,t')\delta\boldsymbol{x}^{\rm s}(t') + \boldsymbol{\chi}(t) \tag{B.1}$$

as we discussed in Sec. 4. The memory kernel $\boldsymbol{M}^{\rm ss}(t,t')$ and the covariance of the effective noise $\boldsymbol{N}_0^{\rm ss}(t,t') = \langle\boldsymbol{\chi}(t)\boldsymbol{\chi}(t')^T\rangle$ can be derived by direct elimination of the bulk variables from (4.1), which is equivalent to substituting the solution for $\delta\boldsymbol{x}^{\rm b}$,

$$\delta\boldsymbol{x}^{\rm b}(t) = {\rm T}\left[e^{\int_0^t \boldsymbol{K}^{\rm bb}(s)ds}\right]\delta\boldsymbol{x}^{\rm b}(0) + \int_0^t {\rm T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\rm bb}(s)ds}\right]\left(\boldsymbol{K}^{\rm bs}(t')\delta\boldsymbol{x}^{\rm s}(t') + \boldsymbol{\xi}^{\rm b}(t')\right)dt' \tag{B.2}$$

(where T is the time-ordering operator). In this way one straightforwardly finds the explicit expression for the effective dynamics (B.1). One reads off that the memory function is given by (4.5), namely

$$\boldsymbol{M}^{\rm ss}(t,t') = \boldsymbol{K}^{\rm sb}(t)\,{\rm T}\left[e^{\int_{t'}^t \boldsymbol{K}^{\rm bb}(s)ds}\right]\boldsymbol{K}^{\rm bs}(t') \tag{B.3}$$

The effective noise covariance has the expression

$$\begin{aligned}\boldsymbol{N}_0^{\rm ss}(t,t') &= \boldsymbol{K}^{\rm sb}(t)\,{\rm T}\left[e^{\int_0^t \boldsymbol{K}^{\rm bb}(s)ds}\right]\boldsymbol{C}^{\rm bb}(0,0)\,{\rm T}\left[e^{\int_0^{t'}(\boldsymbol{K}^{\rm bb}(s))^T ds}\right](\boldsymbol{K}^{\rm sb})^T(t') + \langle\boldsymbol{\Sigma}^{\rm ss}(\boldsymbol{x}(t))\rangle\delta(t-t') \\ &+ \boldsymbol{K}^{\rm sb}(t)\int_0^{\min(t,t')} dt''\,{\rm T}\left[e^{\int_{t''}^t \boldsymbol{K}^{\rm bb}(s)ds}\right]\langle\boldsymbol{\Sigma}^{\rm bb}(\boldsymbol{x}(t))\rangle\,{\rm T}\left[e^{\int_{t''}^{t'}(\boldsymbol{K}^{\rm bb})^T(s)ds}\right](\boldsymbol{K}^{\rm sb})^T(t'')\end{aligned} \tag{B.4}$$

which can be also rewritten as

$$\begin{aligned}\boldsymbol{N}_0^{\rm ss}(t,t') =& \boldsymbol{K}^{\rm sb}(t)\boldsymbol{R}^{\rm bb}(t)\boldsymbol{C}^{\rm bb}(0,0)\boldsymbol{R}^{\rm bb\,T}(t')(\boldsymbol{K}^{\rm sb})^T(t') + \langle\boldsymbol{\Sigma}^{\rm ss}(\boldsymbol{x}(t))\rangle\delta(t-t') \\ &+ \boldsymbol{K}^{\rm sb}(t)\int_0^{\min(t,t')} dt''\boldsymbol{R}^{\rm bb}(t,t'')\langle\boldsymbol{\Sigma}^{\rm bb}(\boldsymbol{x}(t))\rangle\boldsymbol{R}^{\rm bb\,T}(t',t'')(\boldsymbol{K}^{\rm sb})^T(t'')\end{aligned} \tag{B.5}$$

by using the definition (A.30) for the marginal bulk responses.

Appendix B.2. Stationary case

If we consider a dynamics linearized around stationary means (with the diffusion matrix and the dynamical matrix both evaluated then at steady state, i.e. $\langle\boldsymbol{\Sigma}(\boldsymbol{x}(t))\rangle \equiv \boldsymbol{\Sigma}$ and $\boldsymbol{K}(t) \equiv \boldsymbol{K}$), the memory (B.3) becomes

$$\boldsymbol{M}^{\rm ss}(t,t') = \boldsymbol{K}^{\rm sb}e^{\boldsymbol{K}^{\rm bb}(t-t')}\boldsymbol{K}^{\rm bs} \tag{B.6}$$

(precisely expression (4.6)), while the covariance (B.5) of the noise in the subnetwork simplifies to

$$\begin{aligned}\boldsymbol{N}_0^{\rm ss}(t,t') =& \boldsymbol{\Sigma}^{\rm ss}\delta(t-t') + \boldsymbol{K}^{\rm sb}e^{\boldsymbol{K}^{\rm bb}t}\boldsymbol{C}^{\rm bb}(0,0)e^{(\boldsymbol{K}^{\rm bb})^T t'}(\boldsymbol{K}^{\rm sb})^T \\ &+ \boldsymbol{K}^{\rm sb}\int_0^{\min(t,t')} dt''e^{\boldsymbol{K}^{\rm bb}(t-t'')}\boldsymbol{\Sigma}^{\rm bb}e^{(\boldsymbol{K}^{\rm bb})^T(t'-t'')}(\boldsymbol{K}^{\rm sb})^T\end{aligned} \tag{B.7}$$

Appendix C. Nonlinear corrections by perturbation theory

We give details here of the effective action including the first order nonlinear corrections to the Gaussian approximation. The effective action can be treated as a field theoretic free energy and can be expressed perturbatively [64]. Let us decompose the action into a "non-interacting" part (containing only terms quadratic in all the variables) and an interacting one (containing higher powers), which can be seen as a small perturbation

$$\mathcal{H} = \mathcal{H}_0 + \Delta\mathcal{H} \tag{C.1}$$

The expression for the effective action (5.1), applying the idea of a perturbative expansion, becomes

$$\begin{aligned}\mathcal{H}_{\text{eff}} &= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0+\Delta\mathcal{H}} = \\ &= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} \left(1 + \frac{\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}}{\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}} \Delta\mathcal{H} + \mathcal{O}(\Delta\mathcal{H})^2\right) = \\ &= \ln \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} \left(1 + \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}}) \Delta\mathcal{H} + \mathcal{O}(\Delta\mathcal{H})^2\right)\end{aligned} \tag{C.2}$$

where $e^{\mathcal{H}_0} = Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}}, \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$ and $\int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0} = Q_0(\boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$, therefore $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}}) = e^{\mathcal{H}_0} / \int D\boldsymbol{x}^{\text{b}} D\hat{\boldsymbol{x}}^{\text{b}} e^{\mathcal{H}_0}$. The reference distribution for the perturbative expansion $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$ is chosen as the LNA, i.e. a Gaussian conditional distribution centered around the deterministic steady states. To define the effective action we fix subnetwork variables, i.e. condition on them, and need the corresponding conditional bulk probabilities. The perturbative corrections arise from cubic terms in the action, which explicitly read

$$\begin{aligned}\Delta\mathcal{H} = \int_0^T dt \Bigg\{ \sum_i \mathrm{i}\hat{x}_i(t) \Big[\sum_{j,l,j\neq l} k^+_{ij,l} \delta x_i(t) \delta x_j(t) - \frac{1}{2} \sum_{j,l,j\neq l} k^+_{jl,i} \delta x_j(t) \delta x_l(t) + \\ \sum_l k^+_{ii,l} \delta x_i(t) \delta x_i(t) - \frac{1}{2} \sum_j k^+_{jj,i} \delta x_j(t) \delta x_j(t) \Big] \Bigg\}\end{aligned} \tag{C.3}$$

The indices can belong to bulk or subnetwork, giving different combinations; we introduce a vectorial notation

$$\begin{aligned}-\Delta\mathcal{H} = \int_0^T dt \Big\{ &\mathrm{i}\hat{\boldsymbol{x}}^{\text{s}\,T} \Big[\boldsymbol{K}^{\text{s,ss}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{s}}\big) + \boldsymbol{K}^{\text{s,sb}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{b}}\big) + \boldsymbol{K}^{\text{s,bb}}\big(\delta\boldsymbol{x}^{\text{b}} \circ \delta\boldsymbol{x}^{\text{b}}\big) \Big] + \\ &\mathrm{i}\hat{\boldsymbol{x}}^{\text{b}\,T} \Big[\boldsymbol{K}^{\text{b,ss}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{s}}\big) + \boldsymbol{K}^{\text{b,sb}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{b}}\big) + \boldsymbol{K}^{\text{b,bb}}\big(\delta\boldsymbol{x}^{\text{b}} \circ \delta\boldsymbol{x}^{\text{b}}\big) \Big] \Big\}\end{aligned} \tag{C.4}$$

where $\circ$ indicates a "flattened" outer product (see main text) and $\boldsymbol{K}^{\text{s,sb}}$, $\boldsymbol{K}^{\text{b,sb}}$, $\boldsymbol{K}^{\text{b,bb}}, \boldsymbol{K}^{\text{b,ss}}, \boldsymbol{K}^{\text{s,bb}}$ are matrices collecting the relevant nonlinear couplings. As explained in the main text, we do not consider all the possible reactions occurring between subnetwork and bulk, thus we set $\boldsymbol{K}^{\text{b,ss}} = \boldsymbol{K}^{\text{s,bb}} \equiv 0$ (though these would be easy to re-instate). We have also already dropped δ from $\hat{\boldsymbol{x}}^{\text{s}}$ and $\hat{\boldsymbol{x}}^{\text{b}}$ as the marginal auxiliary means $\hat{\boldsymbol{\mu}}^{\text{s}} = \hat{\boldsymbol{\mu}}^{\text{b}} \equiv 0$.

The first order correction (as clear from the second term in (C.2)) reduces to an average w.r.t. the conditional distribution $Q_0(\boldsymbol{x}^{\text{b}}, \hat{\boldsymbol{x}}^{\text{b}} | \boldsymbol{x}^{\text{s}}, \hat{\boldsymbol{x}}^{\text{s}})$. This average is equivalent to replacing combinations of bulk variables in (C.3) by conditional Gaussian moments which, by Wick's theorem, can be expressed in terms of means and correlations

$$\begin{aligned}-\langle\Delta\mathcal{H}\rangle = \int_0^T dt \Big\{ &\mathrm{i}\hat{\boldsymbol{x}}^{\text{s}\,T} \Big[\boldsymbol{K}^{\text{s,ss}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{x}^{\text{s}}\big) + \boldsymbol{K}^{\text{s,sb}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{\mu}^{\text{b|s}}\big) \Big] + \\ &\mathrm{i}\hat{\boldsymbol{\mu}}^{\text{b|s}\,T} \Big[\boldsymbol{K}^{\text{b,sb}}\big(\delta\boldsymbol{x}^{\text{s}} \circ \delta\boldsymbol{\mu}^{\text{b|s}}\big) + \boldsymbol{K}^{\text{b,bb}}\big(\delta\boldsymbol{\mu}^{\text{b|s}} \circ \delta\boldsymbol{\mu}^{\text{b|s}} + \boldsymbol{C}^{\text{bb|s}}(t,t)\big) \Big] \Big\}\end{aligned} \tag{C.5}$$

where one uses that the conditional average $\langle\delta\boldsymbol{x}^{\text{b}}(t)\delta\boldsymbol{x}^{\text{b}\,T}(t)\rangle = \delta\boldsymbol{\mu}^{\text{b|s}}(t)\delta\boldsymbol{\mu}^{\text{b|s}\,T}(t) + \boldsymbol{C}^{\text{bb|s}}(t,t)$. In developing the Wick's theorem for moments of the schematic form $\langle\hat{\boldsymbol{x}}^{\text{b}}\delta\boldsymbol{x}^{\text{b}}\delta\boldsymbol{x}^{\text{b}}\rangle$ or $\langle\hat{\boldsymbol{x}}^{\text{b}}\delta\boldsymbol{x}^{\text{b}}\delta\boldsymbol{x}^{\text{s}}\rangle$ we have used that the

conditional equal time response functions $\boldsymbol{R}^{\mathrm{bb|s}}(t,t)=0$ vanish. We next substitute the expression for $\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}$ (5.7)

$$-\langle\Delta\mathcal{H}\rangle=\int_0^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\Big[\boldsymbol{K}^{\mathrm{s,ss}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{x}^{\mathrm{s}}(t)\big)+\boldsymbol{K}^{\mathrm{s,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\big)\Big]+ \tag{C.6}$$
$$\int_0^T dt'\int_{t'}^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\Big[\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t')\big)+\boldsymbol{K}^{\mathrm{b,bb}}\big(\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t')\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t')+\boldsymbol{C}^{\mathrm{bb|s}}(t',t')\big)\Big]$$

We have re-instated the explicit time dependences here, which are important for our derivation of the memory functions. One can extract the nonlinear reduced dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ from this action by considering the terms multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)$ ($\boldsymbol{C}^{\mathrm{bb|s}}(t',t')$ gives an additional deterministic contribution that we will analyze in Appendix E). First, we note that $\boldsymbol{K}^{\mathrm{s,ss}}$ appearing in the first line of (C.6) gives the nonlinear subnetwork dynamics, which is local in time. The reduced dynamics for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ includes a memory $\boldsymbol{M}^{\mathrm{ss}}(t,t')$ and a coloured noise $\boldsymbol{\chi}(t)$ with covariance

$$\langle\boldsymbol{\chi}(t)\boldsymbol{\chi}^T(t')\rangle=\boldsymbol{N}^{\mathrm{ss}}(t,t') \tag{C.7}$$

to be determined. One can write $\boldsymbol{N}^{\mathrm{ss}}(t,t')=\boldsymbol{N}_0^{\mathrm{ss}}(t,t')+\boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ as the sum of a purely Gaussian term and a first order correction. Similarly the memory comprises two terms: $\boldsymbol{M}^{\mathrm{ss}}(t-t')$, which like $\boldsymbol{N}_0^{\mathrm{ss}}(t,t')$ can be calculated starting from the quadratic part of the action (as shown in Appendix B) and $\boldsymbol{M}^{\mathrm{s,ss}}(t,t',t'')$, which like $\boldsymbol{N}_1^{\mathrm{ss}}(t,t')$ contains the contributions of cubic terms treated by means of perturbation theory. To evaluate these first order corrections, we insert in (C.6) the expressions for $\boldsymbol{\mu}^{\mathrm{b|s}}(t)$ (5.9): to illustrate the procedure we focus on the first term in the second line of (C.6), which becomes

$$\int_0^T dt'\int_{t'}^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t')\big)= \tag{C.8}$$
$$\int_0^T dt'\int_{t'}^T dt\,\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\Big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\int_0^{t'}dt''e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')$$
$$-\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\int_0^T dt''\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'')\Big)$$

The first of the resulting two terms contributes to the reduced subnetwork dynamics via a temporal integral defining a nonlinear memory term given by

$$\int_0^t dt'\int_0^{t'}dt''\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ e^{\boldsymbol{K}^{\mathrm{bb}}(t'-t'')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t'')\big) \tag{C.9}$$

This describes how *products* of subnetwork concentrations feed back into the evolution of a single concentration. More generally, one can see that terms $\sim\hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$ define the nonlinear memory function. Extracting all of these terms from (C.6) gives the result (5.12) stated in the main text (where the kernel of (C.9) appears in the second line). In the second term of (C.8), the factor multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}\,T}(t)$ and $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t'')$

$$\int_0^t dt'\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T\big) \tag{C.10}$$

contributes to the covariance of the effective noise (which always enters the MSRJD formalism via quadratic terms in the auxiliary variables, as is clear from (2.9)) with an additional $\boldsymbol{x}^{\mathrm{s}}$-dependence. Treating the other terms in (C.6) in the same way and dropping third order terms $\sim\hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}$ that encode non-vanishing higher cumulants of the noise distribution one obtains the perturbative contribution to the effective noise covariance, given by all terms of the form $\sim\hat{\boldsymbol{x}}^{\mathrm{s}}\hat{\boldsymbol{x}}^{\mathrm{s}}\delta\boldsymbol{x}^{\mathrm{s}}$; explicitly

$$\boldsymbol{N}_1^{\mathrm{ss}}(t,t'')=\boldsymbol{K}^{\mathrm{s,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t)\circ\boldsymbol{C}^{\mathrm{bb|s}}(t,t'')(\boldsymbol{K}^{\mathrm{sb}})^T\big)+\int_0^t dt'\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-t')}\boldsymbol{K}^{\mathrm{b,sb}}\big(\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(t',t'')(\boldsymbol{K}^{\mathrm{sb}})^T\big)$$
$$+\int_0^t ds\int_0^s dt'\boldsymbol{K}^{\mathrm{sb}}e^{\boldsymbol{K}^{\mathrm{bb}}(t-s)}\Big[\boldsymbol{K}^{\mathrm{b,bb}}\big(e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')\circ\boldsymbol{C}^{\mathrm{bb|s}}(s,t'')(\boldsymbol{K}^{\mathrm{sb}})^T\big)+ \tag{C.11}$$
$$\boldsymbol{K}^{\mathrm{b,bb}}\big(\boldsymbol{C}^{\mathrm{bb|s}}(s,t'')(\boldsymbol{K}^{\mathrm{sb}})^T\circ e^{\boldsymbol{K}^{\mathrm{bb}}(s-t')}\boldsymbol{K}^{\mathrm{bs}}\delta\boldsymbol{x}^{\mathrm{s}}(t')\big)\Big]+\ldots^T$$

where $\ldots^T$ indicates that to each written term its transpose has to be added, as is required to make $\boldsymbol{N}_1^{\rm ss}(t,t'')$ symmetric. We have extended our circle product notation here to products of vectors and matrices, which are defined with the same ordering restrictions as the original product. For example, in the second line of (C.11) we have a term of the form $\boldsymbol{K}^{\rm b,bb}(\boldsymbol{v}^{\rm b}\circ\boldsymbol{A}^{\rm bs})$ where $\boldsymbol{v}^{\rm b}$ is a vector and $\boldsymbol{A}^{\rm bs}$ a matrix. This term is to be read as a matrix with ij-element $\sum_{k\leq l}K_{i,kl}\big(v_kA_{lj}\big)$. For the sake of completeness, one should also include corrections stemming from the fact that the white noise covariance $\boldsymbol{\Sigma}(\boldsymbol{x})$ as given by (2.4) *depends* on concentrations. We defer an analysis of these to a dedicate future paper on stochastic effects in subnetwork modelling.

Appendix D. Effective equations solver

It turns out that the numerical solution of the *integro-differential* equations (5.11) can be simplified by mapping them to a system of ordinary *differential* equations. The idea here is that every memory integral term can be seen as the solution of a differential equation. Let us start from the subnetwork reduced dynamics in the form

$$\frac{d\delta\boldsymbol{x}^{\rm s}(t)}{dt}=\boldsymbol{K}^{\rm ss}\delta\boldsymbol{x}^{\rm s}(t)+\boldsymbol{K}^{\rm s,ss}\big(\delta\boldsymbol{x}^{\rm s}(t)\circ\delta\boldsymbol{x}^{\rm s}(t)\big)+\mathcal{M}(t) \tag{D.1}$$

where we have taken the large volume limit in order to neglect the intrinsic noise. By stepping back through the derivation of the memory vector $\mathcal{M}(t)$ in terms of bulk conditional means $\delta\boldsymbol{\mu}^{\rm b|s}(t)=\boldsymbol{\nu}(t)+\hat{\boldsymbol{\nu}}(t)$, we can rewrite it as in (5.13), i.e.

$$\begin{aligned}\mathcal{M}(t)=&\boldsymbol{K}^{\rm sb}\boldsymbol{\nu}(t)+\boldsymbol{K}^{\rm s,sb}\big(\delta\boldsymbol{x}^{\rm s}(t)\circ\boldsymbol{\nu}(t)\big)+\\&\int_0^t dt'\boldsymbol{K}^{\rm sb}e^{\boldsymbol{K}^{\rm bb}(t-t')}\left[\boldsymbol{K}^{\rm b,sb}\big(\delta\boldsymbol{x}^{\rm s}(t')\circ\boldsymbol{\nu}(t')\big)+\boldsymbol{K}^{\rm b,bb}\big(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t')\big)\right]\end{aligned} \tag{D.2}$$

with $\boldsymbol{\nu}(t')=\int_0^{t'}dt''\big[e^{\boldsymbol{K}^{\rm bb}(t'-t'')}\boldsymbol{K}^{\rm bs}\big]\delta\boldsymbol{x}^{\rm s}(t'')$, a vector of size $N^{\rm b}$ (number of bulk species). The $\boldsymbol{\nu}(t)$ can be obtained equivalently by solving the $N^{\rm b}$ *linear* differential equations

$$\frac{d\boldsymbol{\nu}(t)}{dt}=\boldsymbol{K}^{\rm bb}\boldsymbol{\nu}(t)+\boldsymbol{K}^{\rm bs}\delta\boldsymbol{x}^{\rm s}(t) \tag{D.3}$$

The time integral left in the last term of (D.2) appears as a result of the insertion of $\hat{\boldsymbol{\mu}}^{\rm b|s}(t)$ in the effective action. This term can be rearranged and evaluated by introducing *ad hoc* auxiliary variables. We need to decompose the exponential kernel $e^{\boldsymbol{K}^{\rm bb}(t-t')}$ into a superposition of pure exponentials (by diagonalizing $\boldsymbol{K}^{\rm bb}$)

$$\begin{aligned}&\int_0^t dt'\boldsymbol{K}^{\rm sb}e^{\boldsymbol{K}^{\rm bb}(t-t')}\left[\boldsymbol{K}^{\rm b,sb}\big(\delta\boldsymbol{x}^{\rm s}(t')\circ\boldsymbol{\nu}(t')\big)+\boldsymbol{K}^{\rm b,bb}\big(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t')\big)\right]=\\&=\int_0^t dt'\boldsymbol{K}^{\rm sb}\sum_{c=1}^{N^{\rm b}}\boldsymbol{r}^c\boldsymbol{l}^{c\,T}e^{\lambda_c(t-t')}\left[\boldsymbol{K}^{\rm b,sb}\big(\delta\boldsymbol{x}^{\rm s}(t')\circ\boldsymbol{\nu}(t')\big)+\boldsymbol{K}^{\rm b,bb}\big(\boldsymbol{\nu}(t')\circ\boldsymbol{\nu}(t')\big)\right]\end{aligned} \tag{D.4}$$

where the λ_c are the $N^{\rm b}$ eigenvalues of $\boldsymbol{K}^{\rm bb}$ (with negative real part to ensure stability), $\boldsymbol{r}^c$ and $\boldsymbol{l}^c$ are respectively the right and left eigenvectors of $\boldsymbol{K}^{\rm bb}$, playing the role of the coefficients of this decomposition. We consider a vector $\boldsymbol{z}(t)=\{z^c(t)\}$, $c=1,...,N^{\rm b}$ whose components, one for each eigenvalue λ_c, satisfy

$$\frac{dz^c(t)}{dt}=\lambda_c\,z^c(t)+\boldsymbol{l}^{c\,T}\left[\boldsymbol{K}^{\rm b,sb}\big(\delta\boldsymbol{x}^{\rm s}(t)\circ\boldsymbol{\nu}(t)\big)+\boldsymbol{K}^{\rm b,bb}\big(\boldsymbol{\nu}(t)\circ\boldsymbol{\nu}(t)\big)\right]\qquad z^c(0)=0 \tag{D.5}$$

Then (D.1) can be translated into

$$\frac{d\delta\boldsymbol{x}^{\rm s}(t)}{dt}=\boldsymbol{K}^{\rm ss}\delta\boldsymbol{x}^{\rm s}(t)+\boldsymbol{K}^{\rm s,ss}\big(\delta\boldsymbol{x}^{\rm s}(t)\circ\delta\boldsymbol{x}^{\rm s}(t)\big)+\boldsymbol{K}^{\rm sb}\boldsymbol{\nu}(t)+\boldsymbol{K}^{\rm s,sb}\big(\delta\boldsymbol{x}^{\rm s}(t)\circ\boldsymbol{\nu}(t)\big)+\boldsymbol{K}^{\rm sb}\sum_{c=1}^{N^{\rm b}}\boldsymbol{r}^cz^c(t) \tag{D.6}$$

To summarize, solving the N^{s} subnetwork equations with integral memory terms is equivalent to solving a system with $2N^{\mathrm{b}}$ additional equations, the N^{b} ones describing $\boldsymbol{\nu}(t)$ (see (D.3)) and the N^{b} ones for $\boldsymbol{z}(t)$ (see (D.5)). For projection methods, the additional variables one needs to introduce in order to express memory integrals via differential equations is given by the dimension of the bulk variable space (see [26]) including all the possible concentration products, and hence scales as $(N^{\mathrm{b}} \times N^{\mathrm{b}}) + (N^{\mathrm{s}} \times N^{\mathrm{b}})$.

Appendix E. Projection methods vs GVA

The application of projection methods to protein-protein interaction networks has been studied by Rubin et al. in [26]: here we summarize the basic aspects needed to develop a comparison with the GVA.

One is interested in deriving a closed-form expression for the memory function and the random force both for the linearized and the fully nonlinear dynamics: in the second case, one has to appeal to the limit of vanishing noise. As the variance of copy number fluctuations scales as $\epsilon = 1/V$, i.e. the inverse of the reaction volume, the contribution of the noise becomes negligible for $\epsilon \to 0$, i.e. for suitably large reaction volumes (see Sec. 2 in the main text). While one needs a nonzero ϵ for initially applying the Zwanzig-Mori formalism, the authors always then take the small ϵ limit. Let us stress that the small ϵ limit is not necessary for the linear dynamics, as the noise drops out from the equations for conditionally averaged concentrations whatever the value of ϵ.

In addition, to evaluate the memory and the random force from projection operators, one needs the steady state distribution for the deviations $\delta\boldsymbol{x}$. If the noise is negligible ($\epsilon \to 0$), $\delta\boldsymbol{x}$ is small and one can find the steady state distribution by linearizing the Langevin equation around $\delta\boldsymbol{x} = 0$; this gives a Gaussian steady state distribution of $\delta\boldsymbol{x}$ with covariance satisfying the Lyapunov equation. The structure of the covariance is not unique, but fixed by the type of fluctuations. The choice of independent Poisson fluctuations for each species, thus a covariance with a diagonal structure, produces the simplest projected equations. These describe the evolution for the subnetwork conditionally on the available knowledge of the initial conditions which are specified, for the bulk, via some probability distribution. The solution is the mean trajectory w.r.t. to this initial distribution, from which stochastic fluctuations are thus averaged out: we will use for it the same notation as for the single instances $\boldsymbol{x}^{\mathrm{s}}(t)$ as they coincide in the limit $\epsilon \to 0$. As a consequence, we expect that the equations of motion given by projection methods for such conditionally averaged concentrations agree with the noise averaged subnetwork equations in the GVA. We next look closely at this comparison.

Appendix E.1. Linearized projected equations

The starting point is the Langevin dynamics (2.1) and its corresponding Fokker-Planck equation, encoded by $\mathcal{L}$, the so-called adjoint Fokker-Planck operator. Referring to [26] for the entire derivation, we shall directly provide the final form of the linearized dynamics

$$\frac{d\delta x_i(t)}{dt} = \sum_{j=1}^{N^{\mathrm{s}}} \delta x_j(t) \boldsymbol{\Omega}^{\mathrm{ss}}_{ji} + \sum_{j=1}^{N^{\mathrm{s}}} \int_0^t dt' \delta x_j(t') \boldsymbol{M}^{\mathrm{ss}}_{ji}(t-t') + r_i(t) \tag{E.1}$$

This is obtained by applying two operators projecting either onto the subspace of subnetwork d.o.f. (degrees of freedom) or onto the orthogonal one; in the linearized dynamics, it is found that they can be represented by matrices with a simple block structure, $\boldsymbol{P}$ and $\boldsymbol{Q}$ respectively

$$\boldsymbol{P} = \begin{pmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{pmatrix}$$

$$\boldsymbol{Q} = \begin{pmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{pmatrix}$$

The adjoint Fokker-Planck operator $\mathcal{L}$ can also be cast in matrix form and is then denoted as $\boldsymbol{L}$. Note that $\boldsymbol{L} = \boldsymbol{K}^T$, i.e. it is equivalent to the transpose of dynamical matrix of the GVA; like $\boldsymbol{K}$ it can be partitioned

into blocks for the bulk and the subnetwork part as

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{\mathrm{ss}} & \boldsymbol{L}^{\mathrm{sb}} \\ \boldsymbol{L}^{\mathrm{bs}} & \boldsymbol{L}^{\mathrm{bb}} \end{pmatrix}$$

Exploiting this correspondence between operators and matrices one has for the linear dynamics:

- $\boldsymbol{\Omega}$ is the top left block (related to subnetwork variables) of $\boldsymbol{PL}$, thus $\boldsymbol{\Omega} = \boldsymbol{L}^{\mathrm{ss}}$
- $\boldsymbol{M}(t-t')$ is the top left block of $\boldsymbol{PLQ}e^{\boldsymbol{QLQ}(t-t')}\boldsymbol{QL}$, i.e.

$$\boldsymbol{M}(t-t') = \boldsymbol{L}^{\mathrm{sb}} e^{\boldsymbol{L}^{\mathrm{bb}}(t-t')} \boldsymbol{L}^{\mathrm{bs}} \tag{E.2}$$

- The random force is given by the s-entries of $\delta \boldsymbol{x}^{\mathrm{b}\,T}(0) e^{\boldsymbol{QLQ}t}\boldsymbol{QL}$, i.e.

$$\boldsymbol{r}_0^T(t) = \delta \boldsymbol{x}^{\mathrm{b}\,T}(0) e^{\boldsymbol{L}^{\mathrm{bb}}t} \boldsymbol{L}^{\mathrm{bs}} \tag{E.3}$$

Note that with the index conventions used in the projection method [26], the memory function above is the matrix *transpose* of the one in our GVA approach. To avoid introducing more notation, we use the same symbol $\boldsymbol{M}$ nonetheless.

Appendix E.2. Nonlinear projected equations

For the nonlinear projected equations one again starts from suitable matrix representations of the operators involved. Note that in the projection method, even once a subnetwork has been selected, there is freedom in choosing the space of observables onto which to project. Rubin et al. show [26] that the best option is to project onto concentrations and products of concentrations.

The nonlinearity is represented as a linear coupling between a concentration and products of two concentrations; the projected equations have the form

$$\begin{aligned} \frac{d\delta x_i(t)}{dt} = &\sum_{j=1}^{N^{\mathrm{s}}} \delta x_j(t) \boldsymbol{\Omega}^{\mathrm{ss}}_{ji} + \sum_{1\le j\le k\le N^{\mathrm{s}}} \delta x_j(t)\delta x_k(t) \boldsymbol{\Omega}^{\mathrm{ss,s}}_{(jk),i} + \int_0^t dt' \Bigg(\sum_{j=1}^{N^{\mathrm{s}}} \delta x_j(t') \boldsymbol{M}^{\mathrm{ss}}_{ji}(t-t') \\ &+ \sum_{1\le j\le k\le N^{\mathrm{s}}} \delta x_j(t')\delta x_k(t') \boldsymbol{M}^{\mathrm{ss,s}}_{(jk),i}(t-t') \Bigg) + r_i(t) \end{aligned} \tag{E.4}$$

Focussing on an observable z_α (summarizing both simple concentrations and products), one can write the adjoint Fokker-Planck operator $\mathcal{L}$ in a matrix form such that

$$\partial_t z_\alpha = \sum_\beta z_\beta L_{\beta\alpha} + \delta x^3 + \mathcal{O}(\epsilon) \tag{E.5}$$

δx^3 represents cubic terms, which are not captured at this order of accuracy, while $\mathcal{O}(\epsilon)$-terms vanish in the small ϵ limit. The matrix representation therefore mirrors the choice of an enlarged space containing also products, and it is valid for small ϵ. It reads explicitly as follows

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^{\mathrm{ss}} & \boldsymbol{L}^{\mathrm{sb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{bs}} & \boldsymbol{L}^{\mathrm{bb}} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \boldsymbol{L}^{\mathrm{ss,s}} & \boldsymbol{L}^{\mathrm{ss,b}} & \boldsymbol{L}^{\mathrm{ss,ss}} & \boldsymbol{L}^{\mathrm{ss,sb}} & \boldsymbol{L}^{\mathrm{ss,bb}} \\ \boldsymbol{L}^{\mathrm{sb,s}} & \boldsymbol{L}^{\mathrm{sb,b}} & \boldsymbol{L}^{\mathrm{sb,ss}} & \boldsymbol{L}^{\mathrm{sb,sb}} & \boldsymbol{L}^{\mathrm{sb,bb}} \\ \boldsymbol{L}^{\mathrm{bb,s}} & \boldsymbol{L}^{\mathrm{bb,b}} & \boldsymbol{L}^{\mathrm{bb,ss}} & \boldsymbol{L}^{\mathrm{bb,sb}} & \boldsymbol{L}^{\mathrm{bb,bb}} \end{pmatrix}$$

$\boldsymbol{L}$ consists of 5 block rows and columns, referring to linear subnetwork concentrations (s), subnetwork products (ss), mixed subnetwork-bulk products (sb), which are considered part of the bulk subspace, and products of bulk concentrations (bb). The dynamics for the products, contained in the bottom right blocks ($\boldsymbol{L}^{\mathrm{ss,ss}}$, $\boldsymbol{L}^{\mathrm{ss,sb}}$ etc.), is simply derived from linearized dynamics, thus

$$\partial_t(\delta x_i \delta x_j) = \delta x_j \partial_t \delta x_i + \delta x_i \partial_t \delta x_j \tag{E.6}$$

In other words, in the evolution of a product only products appear: this explains why the top right block vanishes, i.e. applying $\boldsymbol{L}$ to quadratic observables does not give linear terms. All the coefficients in the bottom right blocks are simply "imported" from the linearized dynamics. Genuine nonlinearities enter via the bottom left blocks ($\boldsymbol{L}^{\mathrm{ss,s}}$, $\boldsymbol{L}^{\mathrm{ss,b}}$ etc.). These contain the coefficients multiplying products in the equations for linear observables. For the projection matrices one has

$$\boldsymbol{P} = \begin{pmatrix} \mathbb{1} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{1} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \end{pmatrix}$$

while $\boldsymbol{Q}$ has the roles of $\mathbb{1}$ and $\mathbb{0}$ along the diagonal are interchanged

$$\boldsymbol{Q} = \begin{pmatrix} \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{1} \end{pmatrix}$$

The combinations of matrices one needs, to generalize the formulas of the previous section, are $\boldsymbol{PL}$, $\boldsymbol{QLQ}$ and $\boldsymbol{QL}$, see [26] for explicit expressions. We emphasize that $\boldsymbol{QLQ}$ has a lower triangular block structure, thus $\boldsymbol{E}(t) = e^{\boldsymbol{QLQ}t}$ has the same structure with diagonal blocks that are the exponentials of those in $\boldsymbol{QLQ}$, i.e. for example $[e^{\boldsymbol{QLQ}t}]_{\mathrm{bb}} = e^{\boldsymbol{L}^{\mathrm{bb}}t}$. Another property that can be deduced from such a structure is

$$\boldsymbol{E}_{\mathrm{sb,b}}(t) = \int_0^t dt' \boldsymbol{E}_{\mathrm{sb,sb}}(t-t')\boldsymbol{L}^{\mathrm{sb,b}}\boldsymbol{E}_{\mathrm{bb}}(t') \tag{E.7a}$$

$$\boldsymbol{E}_{\mathrm{bb,b}}(t) = \int_0^t dt' \boldsymbol{E}_{\mathrm{bb,bb}}(t-t')\boldsymbol{L}^{\mathrm{bb,b}}\boldsymbol{E}_{\mathrm{bb}}(t') \tag{E.7b}$$

The rate constants, the memory function and the random force for the full nonlinear dynamics can now be found in analogy with the linear case provided that we consider the above enlarged matrices; therefore:

- The nonlinear rate matrix for the internal subnetwork dynamics is $\boldsymbol{\Omega}^{\mathrm{ss,s}} = \boldsymbol{L}^{\mathrm{ss,s}}$
- The memory is the ss block of $\boldsymbol{PLQ}e^{\boldsymbol{QLQ}t}\boldsymbol{QL}$, whose nonlinear part is

$$\boldsymbol{M}^{\mathrm{ss,s}}(t-t') = \boldsymbol{L}^{\mathrm{ss,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\cdot\mathrm{b,\cdot b}}(t-t')\boldsymbol{L}^{\mathrm{\cdot b,s}} \tag{E.8}$$

  where "·b" indicates the union of the "sb" and "bb" product ranges. Assumptions about what reactions between the bulk and the subnetwork can occur imply that some $\boldsymbol{L}$ blocks are zero, namely: $\boldsymbol{L}^{\mathrm{ss,bb}} = \boldsymbol{L}^{\mathrm{bb,ss}} = \boldsymbol{L}^{\mathrm{ss,b}} = \boldsymbol{L}^{\mathrm{bb,s}} = \mathbb{0}$ [26]. These constraints then simplify the expression for the memory considerably

$$\boldsymbol{M}^{\mathrm{ss,s}}(t-t') = \boldsymbol{L}^{\mathrm{ss,sb}}\boldsymbol{E}_{\mathrm{sb,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{L}^{\mathrm{ss,sb}}\boldsymbol{E}_{\mathrm{sb,sb}}(t-t')\boldsymbol{L}^{\mathrm{sb,s}} \tag{E.9}$$

  We shall nevertheless stick to the most general case, to keep our arguments general. The memory function in the dynamics for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ is then embedded in a time integral over the past history of all possible subnetwork products $\delta\boldsymbol{x}^{\mathrm{ss}}(t')$, thus the general nonlinear memory for the projected equations is

$$\boldsymbol{\mathcal{M}}^T_{\mathrm{proj}}(t) = m_1 + m_2 + m_3 \tag{E.10}$$

  with

$$m_1 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} \tag{E.11}$$

$$m_2 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} \tag{E.12}$$

$$m_3 = \int_0^t dt' \delta\boldsymbol{x}^{\mathrm{ss}\,T}(t')\boldsymbol{L}^{\mathrm{ss,\cdot b}}\boldsymbol{E}_{\cdot\mathrm{b,\cdot b}}(t-t')\boldsymbol{L}^{\mathrm{\cdot b,s}} \tag{E.13}$$

- The random force cannot be calculated in closed form from the matrix representations introduced, but an expansion can still be provided up to quadratic order in deviations form the steady state: one may hope that higher order terms not captured in this way are small or negligible. We shall write $\boldsymbol{r}(t) = \boldsymbol{r}_0(t) + \boldsymbol{r}_1(t)$, $\boldsymbol{r}_1(t)$ being the nonlinear random force given by

$$\boldsymbol{r}_1^T(t) = \delta\boldsymbol{x}^{\mathrm{sb}\,T}(0)\left[\boldsymbol{E}_{\mathrm{sb,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\mathrm{sb,sb}}(t)\boldsymbol{L}^{\mathrm{sb,s}}\right] + \delta\boldsymbol{x}^{\mathrm{bb}\,T}(0)\left[\boldsymbol{E}_{\mathrm{bb,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\mathrm{bb,sb}}(t)\boldsymbol{L}^{\mathrm{sb,s}}\right] \tag{E.14}$$

From these results, we see that the projection method divides the non-Markovian contribution from the bulk into the random force, which depends only on initial conditions, and the memory, containing all the single points in the past for single species and products. This structural feature is important for the comparison to the GVA, which does not separate the non-local-in-time terms so neatly but still gives an *equivalent* approximation *up to the second order and in the limit* $\epsilon \to 0$. To provide the tools for this comparison, we first recall the perturbative expansion allowing us to derive nonlinear corrections in the Gaussian variational approach (see Appendix E.3) and then explicitly derive this equivalence in Appendix E.4.

Appendix E.3. Full nonlinear memory and random force in the GVA

As we mentioned, the GVA and the projection method are expected to agree only for $\epsilon \to 0$: we then restrict ourselves to this case, i.e. the deterministic dynamics. Let us translate the perturbative approach of Appendix C to a notation analogous to the one in the previous section; in particular the cubic part of the action (C.4) then reads

$$\begin{aligned} -\Delta\mathcal{H} = \int_0^T dt\Bigg\{ &\left[\left(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}\right)^T \boldsymbol{L}^{\mathrm{ss,s}} + \left(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\right)^T \boldsymbol{L}^{\mathrm{sb,s}} + \left(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\right)^T \boldsymbol{L}^{\mathrm{bb,s}}\right] \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}} + \\ &\left[\left(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{s}}\right)^T \boldsymbol{L}^{\mathrm{ss,b}} + \left(\delta\boldsymbol{x}^{\mathrm{s}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\right)^T \boldsymbol{L}^{\mathrm{sb,b}} + \left(\delta\boldsymbol{x}^{\mathrm{b}} \circ \delta\boldsymbol{x}^{\mathrm{b}}\right)^T \boldsymbol{L}^{\mathrm{bb,b}}\right] \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{b}} \Bigg\} \end{aligned} \tag{E.15}$$

In light of the fact that we allow only certain processes, after (C.4) we had invoked the simplification $\boldsymbol{L}^{\mathrm{ss,b}} = \boldsymbol{L}^{\mathrm{bb,s}} \equiv 0$, while here, for the sake of a general comparison, we keep all the nonlinear couplings. The nonlinear memory and effective noise covariance are evaluated by taking the average w.r.t. the Gaussian conditional distribution over bulk variables, as given by (C.6). In the new notation it becomes

$$\begin{aligned} -\langle\Delta\mathcal{H}\rangle = &\int_0^T dt\left(\delta\boldsymbol{x}^{\mathrm{s}}(t) \circ \delta\boldsymbol{x}^{\mathrm{s}}(t)\right)^T \boldsymbol{L}^{\mathrm{ss,s}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \\ &\int_0^T dt\Bigg\{\left[\left(\delta\boldsymbol{x}^{\mathrm{s}}(t) \circ \delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\right)^T \boldsymbol{L}^{\mathrm{sb,s}} + \left(\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) \circ \delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t)\right)^T \boldsymbol{L}^{\mathrm{bb,s}}\right] \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t) + \\ &\int_t^T dt'\left[\left(\delta\boldsymbol{x}^{\mathrm{s}}(t) \circ \delta\boldsymbol{\mu}^{\mathrm{b|s}}(t)\right)^T \boldsymbol{L}^{\mathrm{sb,b}} + \left(\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) \circ \delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) + \boldsymbol{C}^{\mathrm{bb|s}}(t,t)\right)^T \boldsymbol{L}^{\mathrm{bb,b}}\right] \boldsymbol{E}_{\mathrm{bb}}(t'-t)\boldsymbol{L}^{\mathrm{bs}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t')\Bigg\} \end{aligned} \tag{E.16}$$

given that the auxiliary conditional mean, from (5.7), can be rewritten as

$$\mathrm{i}\hat{\boldsymbol{\mu}}^{\mathrm{b|s}}(t) = \int_t^T dt'\, \boldsymbol{E}_{\mathrm{bb}}(t'-t)\boldsymbol{L}^{\mathrm{bs}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{E.17}$$

Then one substitutes the expressions for the conditional means (5.9), which now read

$$\delta\boldsymbol{\mu}^{\mathrm{b|s}}(t) = \boldsymbol{\mu}^{\mathrm{b|s}}(t) - \boldsymbol{\mu}^{\mathrm{b}}(t) = \int_0^t dt' e^{(\boldsymbol{L}^{\mathrm{bb}})^T(t-t')}(\boldsymbol{L}^{\mathrm{sb}})^T \delta\boldsymbol{x}^{\mathrm{s}}(t') - \int_0^T dt' \boldsymbol{C}^{\mathrm{bb|s}}(t,t')\boldsymbol{L}^{\mathrm{bs}} \mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t') \tag{E.18}$$

in (E.16) and by isolating the terms multiplying $\mathrm{i}\hat{\boldsymbol{x}}^{\mathrm{s}}(t)$ one can read the nonlinear reduced dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$. The memory contribution of the GVA comes from the terms that are quadratic in $\delta\boldsymbol{x}^{\mathrm{s}}$ and non-local

in time

$$\boldsymbol{\mathcal{M}}^T_{\rm GVA}(t) = \int_0^t dt' \Big\{ \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t') \circ \delta \boldsymbol{x}^{\rm s\ T}(t)\big) \boldsymbol{L}^{\rm sb,s} + \tag{E.19}$$
$$\int_{t'}^t dt'' \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t''-t') \circ \delta \boldsymbol{x}^{\rm s\ T}(t'')\big) \boldsymbol{L}^{\rm sb,b} \boldsymbol{E}_{\rm bb}(t-t'') \boldsymbol{L}^{\rm bs} +$$
$$\int_0^t dt'' \int_{\max(t',t'')}^t ds \, \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t') \circ \delta \boldsymbol{x}^{\rm s\ T}(t'') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t'')\big) \boldsymbol{L}^{\rm bb,s} +$$
$$\int_0^t dt'' \int_{\max(t',t'')}^t ds \, \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t') \circ \delta \boldsymbol{x}^{\rm s\ T}(t'') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t'')\big) \boldsymbol{L}^{\rm bb,b} \boldsymbol{E}_{\rm bb}(t-s) \boldsymbol{L}^{\rm bs} \Big\}$$

This corresponds to the double time integral of (5.12) in the modified notation and keeping all the couplings. As before, by switching to the projection method notation, we are taking the transposes of the memory functions in the main text, equations (4.6) and (5.12).

Similarly from the terms $\sim \hat{\boldsymbol{x}}^{\rm s} \delta \boldsymbol{x}^{\rm s} \hat{\boldsymbol{x}}^{\rm s}$ we can deduce the effective noise covariance

$$\boldsymbol{N}^{\rm ss}_1(t,t'') = \big((\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(t,t'') \circ \delta \boldsymbol{x}^{\rm s\ T}(t)\big) \boldsymbol{L}^{\rm sb,s} + \int_0^t dt' \big((\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(t',t'') \circ \delta \boldsymbol{x}^{\rm s\ T}(t')\big) \boldsymbol{L}^{\rm sb,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs}$$
$$+ \int_0^t dt' \Big[\big((\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(t,t'') \circ \delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t')\big) + \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t') \circ (\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(t,t'')\big) \Big] \boldsymbol{L}^{\rm bb,s}$$
$$+ \int_0^t ds \int_0^s dt' \Big[\big((\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(s,t'') \circ \delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t')\big) + \tag{E.20}$$
$$\big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t') \circ (\boldsymbol{L}^{\rm bs})^T \boldsymbol{C}^{\rm bb|s}(s,t'')\big) \Big] \boldsymbol{L}^{\rm bb,b} \boldsymbol{E}_{\rm bb}(t-s) \boldsymbol{L}^{\rm bs} + \ldots^T$$

which generalizes (C.11). For $\epsilon \to 0$, the bulk conditional correlator is simply

$$\boldsymbol{C}^{\rm bb|s}(t',t'') = \boldsymbol{E}^T_{\rm bb}(t') \boldsymbol{C}^{\rm bb}(0,0) \boldsymbol{E}_{\rm bb}(t'') \tag{E.21}$$

as can be deduced from (5.10).

Let us stress that projection methods expand the random force, while our perturbative approach expands the correlator of a coloured noise. For the sake of comparison, it is therefore convenient to think in terms of the effective coloured noise with covariance $\boldsymbol{N}^{\rm ss}_0 + \boldsymbol{N}^{\rm ss}_1$. $\boldsymbol{N}^{\rm ss}_1$ can be read from (E.20), while $\boldsymbol{N}^{\rm ss}_0$, the effective noise covariance of the linearized dynamics, is given by (B.7); we rewrite it as

$$\boldsymbol{N}^{\rm ss}_0(t,t') = (\boldsymbol{L}^{\rm bs})^T \boldsymbol{E}^T_{\rm bb}(t) \boldsymbol{C}^{\rm bb}(0,0) \boldsymbol{E}_{\rm bb}(t') \boldsymbol{L}^{\rm bs} \tag{E.22}$$

Note that we have dropped the intrinsic noise contribution $\boldsymbol{\Sigma}^{\rm ss}$, $\boldsymbol{\Sigma}^{\rm bb}$ as we focus on $\epsilon \to 0$. The Gaussian noise $\boldsymbol{\chi}_0(t)$ such that $\langle \boldsymbol{\chi}_0(t) \boldsymbol{\chi}^T_0(t') \rangle = \boldsymbol{N}^{\rm ss}_0(t,t')$ is therefore

$$\boldsymbol{\chi}^T_0(t) = \delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(t) \boldsymbol{L}^{\rm bs} \tag{E.23}$$

If we define $\boldsymbol{\chi}_1(t)$ as follows

$$\boldsymbol{\chi}^T_1(t) = \big(\delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}^{\rm bb}(t) \circ \delta \boldsymbol{x}^{\rm s\ T}(t)\big) \boldsymbol{L}^{\rm sb,s} + \int_0^t dt' \big(\delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(t') \circ \delta \boldsymbol{x}^{\rm s\ T}(t')\big) \boldsymbol{L}^{\rm sb,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs} +$$
$$\int_0^t dt' \Big[\big(\delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(t) \circ \delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t')\big) + \big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t') \circ \delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(t)\big) \Big] \boldsymbol{L}^{\rm bb,s}$$
$$+ \int_0^t ds \int_0^s dt' \Big[\big(\delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(s) \circ \delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t')\big) +$$
$$\big(\delta \boldsymbol{x}^{\rm s\ T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t') \circ \delta \boldsymbol{x}^{\rm b\ T}(0) \boldsymbol{E}_{\rm bb}(s)\big) \Big] \boldsymbol{L}^{\rm bb,b} \boldsymbol{E}_{\rm bb}(t-s) \boldsymbol{L}^{\rm bs} \tag{E.24}$$

then the correlation function of $\boldsymbol{\chi}_0 + \boldsymbol{\chi}_1$ is given by $\boldsymbol{N}_0^{\rm ss} + \boldsymbol{N}_1^{\rm ss}$ at the linear order in $\delta\boldsymbol{x}^{\rm s}$ (this is the order in $\delta\boldsymbol{x}^{\rm s}$ to which $\boldsymbol{N}_1^{\rm ss}$ is calculated). For example, let us take the first term in (E.20): it is derived from cross-correlations of $\boldsymbol{\chi}_0$ (E.23) and the first term of $\boldsymbol{\chi}_1$ (E.24) as follows

$$\left\langle \big((\boldsymbol{L}^{\rm bs})^T \boldsymbol{E}_{\rm bb}^T(t)\delta\boldsymbol{x}^{\rm b}(0)\delta\boldsymbol{x}^{{\rm b}\,T}(0)\boldsymbol{E}_{\rm bb}(t'')\circ\delta\boldsymbol{x}^{{\rm s}\,T}(t'')\big)\boldsymbol{L}^{\rm sb,s}\right\rangle = \big((\boldsymbol{L}^{\rm bs})^T\boldsymbol{C}^{\rm bb|s}(t,t'')\circ\delta\boldsymbol{x}^{{\rm s}\,T}(t'')\big)\boldsymbol{L}^{\rm sb,s} \quad \text{(E.25)}$$

by using the definition (E.21).

From expression (E.16) (see also (C.5)), it can be seen that the Gaussian effective dynamics of $\delta\boldsymbol{x}^{\rm s}(t)$ exhibits also the $\boldsymbol{x}^{\rm s}$-independent term

$$\boldsymbol{C}^{\rm bb|s}(t,t)\boldsymbol{L}^{\rm bb,s} + \int_0^t dt'\boldsymbol{C}^{\rm bb|s}(t',t')\boldsymbol{L}^{\rm bb,b}\boldsymbol{E}_{\rm bb}(t-t')\boldsymbol{L}^{\rm bs} = \langle\boldsymbol{\psi}_1^T(t)\rangle \quad \text{(E.26)}$$

which can be written as the conditional average of a vector $\boldsymbol{\psi}_1(t)$ defined as follows

$$\begin{aligned}\boldsymbol{\psi}_1^T(t) =& \big(\delta\boldsymbol{x}^{{\rm b}\,T}(0)\boldsymbol{E}_{\rm bb}(t)\circ\delta\boldsymbol{x}^{{\rm b}\,T}(0)\boldsymbol{E}_{\rm bb}(t)\big)\boldsymbol{L}^{\rm bb,s} + \\ & \int_0^t dt'\big(\delta\boldsymbol{x}^{{\rm b}\,T}(0)\boldsymbol{E}_{\rm bb}(t')\circ\delta\boldsymbol{x}^{{\rm b}\,T}(0)\boldsymbol{E}_{\rm bb}(t')\big)\boldsymbol{L}^{\rm bb,b}\boldsymbol{E}_{\rm bb}(t-t')\boldsymbol{L}^{\rm bs}\end{aligned} \quad \text{(E.27)}$$

The conditional average of $\boldsymbol{\psi}_1$ produces in the action a cubic term, $\sim \hat{\boldsymbol{x}}^s\delta\boldsymbol{x}^{\rm b}\delta\boldsymbol{x}^{\rm b}$, while its variance would appear in the action via a term of 6th order, thus it is not present at the order we have kept in our calculation. From (E.27) we see that $\boldsymbol{\psi}_1(t)$ is a temporally correlated term in the reduced subnetwork dynamics that depends quadratically on $\delta\boldsymbol{x}^{\rm b}(0)$; it can be regarded as a further contribution to the nonlinear "random force" of the GVA, as by definition the random force summarizes the uncertainty on the bulk initial conditions. As a result, the systematic perturbative expansion up to the cubic order in the action gives it as

$$\tilde{\boldsymbol{r}}_{1\,\rm GVA}(t) = \boldsymbol{\chi}_1(t) + \langle\boldsymbol{\psi}_1(t)\rangle \quad \text{(E.28)}$$

If we consider the fluctuating version of this quantity, namely

$$\boldsymbol{r}_{1\,\rm GVA}(t) = \boldsymbol{\chi}_1(t) + \boldsymbol{\psi}_1(t) \quad \text{(E.29)}$$

we can prove the equivalence with the projection formalism.

Appendix E.4. Proof of the equivalence

Let us assume that the initial deviations from the means are proportional to some factor δ, $\delta\boldsymbol{x}^{\rm s}(0)\sim\delta$ and $\delta\boldsymbol{x}^{\rm b}(0)\sim\delta$. The solution $\delta\boldsymbol{x}^{\rm s}(t)$ can be expanded in powers of δ and we want to find out to what extent projection and GVA give the same results for this expansion in the limit $\epsilon\to 0$. We expect that the two descriptions are equivalent at $\mathcal{O}(\delta^2)$ as in the GVA we keep cubic terms in the effective action while in projection methods we keep all quadratic observables.

Appendix E.4.1. Linear order in δ. To work out the solution of $\delta\boldsymbol{x}^{\rm s}(t)$ in this case, we need to include in its dynamics the expressions for memory and random force at $\mathcal{O}(\delta)$. By re-writing the linear memory (4.6) of the GVA in the notation of Appendix E.2 one has

$$\boldsymbol{M}^{\rm ss}(t-t') = \boldsymbol{L}^{\rm sb}\boldsymbol{E}_{\rm bb}(t-t')\boldsymbol{L}^{\rm bs} \quad \text{(E.30)}$$

where we have used $e^{\boldsymbol{L}^{\rm bb}t} = \boldsymbol{E}_{\rm bb}(t)$. The effective coloured noise at this order for the GVA is given by (E.23). Expressions (E.30) and (E.23) are the same as from projection methods (E.2) and (E.3), thus the resulting expansion of $\delta\boldsymbol{x}^{\rm s}(t)$ is identical to $\mathcal{O}(\delta)$.

Appendix E.4.2. Quadratic order in δ. In this case, we need to include in the dynamics of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ the memory and the effective noise/random force at $\mathcal{O}(\delta^2)$. This includes linear memory terms acting on the second order part of $\delta\boldsymbol{x}^{\mathrm{s}}(t)$, which will be the same for projection and GVA. The remaining terms are the nonlinear memory and nonlinear effective noise/random force evaluated to order $\mathcal{O}(\delta^2)$.

Expressions for nonlinear memory and random force from the two approaches (respectively (E.10) and (E.19), (E.14) and (E.29)) do not coincide if taken *separately*; we nevertheless expect that their *combination* is actually equivalent (because there can be different ways of writing expansions that are correct at the same order, namely $\mathcal{O}(\delta^2)$).

Before verifying this expectation, we briefly outline the reasoning. First of all, the memory function (E.19) differs from the one in the projection methods (E.10): in particular, the two subnetwork species in quadratic terms are calculated at two past times instead of the same time; additional integrals over time also appear. One can however eliminate the intermediate times, and thus match the one-time and two-times structures in the algebra, by explicitly expressing everywhere the dependence on subnetwork initial values. Next we observe that in the structure of the enlarged (i.e. including quadratic observables) $\boldsymbol{L}$, the non-linearity of interactions enters genuinely only in the bottom left off-diagonal block. The bottom right (diagonal) block, which represents the coupling between products, does not provide any additional information with respect to the linear dynamics: it simply describes how the products propagate under the linear dynamics. Note that this is precisely what the perturbative expansion implements, i.e. it describes products under the linear evolution. Therefore, we can drop the purely non-linear block and try to write also the products of subnetwork variables and conditional means of the perturbative expansion as functions of the initial conditions in the subnetwork via the exponential of a matrix consisting of a linear and a quadratic block.

We need to substitute into the nonlinear memory and random force the first order deterministic solution for $\delta\boldsymbol{x}^{\mathrm{s}}(t)$ and $\delta\boldsymbol{x}^{\mathrm{b}}(t)$ to evaluate them consistently to $\mathcal{O}(\delta^2)$: let us perform the substitution in both approaches and compare the results. The projection memory is evaluated at $\mathcal{O}(\delta^2)$ using the expression

$$\delta\boldsymbol{x}^{\cdot\cdot\,T}(t) = \delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t} \tag{E.31}$$

where "$\cdot\cdot$" refers to the block obtained by joining the "ss","sb" and "bb" ranges, here and everywhere below. Also in the GVA we have to write the products as in (E.31) but with the following caveat. The GVA is obtained via the inclusion of the dynamics of conditional *averages*, where the average is taken also over bulk initial conditions. This, importantly, implies that $\delta\boldsymbol{x}^{\mathrm{b}}(0)$ should not be treated as a fluctuating quantity as happens in (E.31). More explicitly, we recall that the conditional bulk variables can be written as $\delta\boldsymbol{\mu}^{\mathrm{b|s}\,T}(t) = \boldsymbol{\nu}^{T}(t) + \hat{\boldsymbol{\nu}}^{T}(t)$, where only the deterministic part $\boldsymbol{\nu}(t)$ contributes to the memory

$$\boldsymbol{\nu}^{T}(t) = \int_0^t dt'\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\,\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{E.32}$$

(from expression (5.9)). On the other hand, the evolution of the products introduced by (E.31) accounts for a solution for bulk variables as follows

$$\delta\boldsymbol{x}^{\mathrm{b}\,T}(t) = \delta\boldsymbol{x}^{\mathrm{b}\,T}(0)\boldsymbol{E}_{\mathrm{bb}}(t) + \int_0^t dt'\delta\boldsymbol{x}^{\mathrm{s}\,T}(t')\boldsymbol{L}^{\mathrm{sb}}\,\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{E.33}$$

with also deviations in the initial conditions $\delta\boldsymbol{x}^{\mathrm{b}}(0)$. The nonlinear memory for the variational approach (E.19) can be written as

$$\begin{aligned}\boldsymbol{\mathcal{M}}^{T}_{\mathrm{GVA}}(t) &= \int_0^t dt'\left\langle\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t'}\right\rangle_{\cdot\cdot}\boldsymbol{L}^{\cdot\cdot,\mathrm{b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \left\langle\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t}\right\rangle_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b,s}} = \\ &= \int_0^t dt'\left(\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t'}\right)_{\cdot\cdot}\boldsymbol{L}^{\cdot\cdot,\mathrm{b}}\boldsymbol{E}_{\mathrm{bb}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \left(\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot\cdot,\cdot\cdot}t}\right)_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b,s}} - m_4' = \\ &= m_1' + m_2' + m_3' - m_4'\end{aligned} \tag{E.34}$$

We have

$$m_1' = \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\rm ss} \boldsymbol{L}^{\rm ss,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs} \tag{E.35}$$

$$m_2' = \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs} \tag{E.36}$$

$$m_3' = (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,s} \tag{E.37}$$

m_4' has minus sign as it is a correction term needed for the substitution in the second line of (E.34) to be valid (in other words, to compensate for the use of (E.33) instead of (E.32)); namely

$$\begin{aligned} m_4' = {} & \big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}^{\rm bb}(t) \circ \delta \boldsymbol{x}^{\rm s\,T}(t)\big) \boldsymbol{L}^{\rm sb,s} + \int_0^t dt' \big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t') \circ \delta \boldsymbol{x}^{\rm s\,T}(t')\big) \boldsymbol{L}^{\rm sb,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs} + \\ & \int_0^t dt' \Big[\big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t) \circ \delta \boldsymbol{x}^{\rm s\,T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t')\big) + \big(\delta \boldsymbol{x}^{\rm s\,T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(t-t') \circ \delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t)\big) \Big] \boldsymbol{L}^{\rm bb,s} + \\ & \big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t) \circ \delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t)\big) \boldsymbol{L}^{\rm bb,s} + \int_0^t dt' \big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t') \circ \delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(t')\big) \boldsymbol{L}^{\rm bb,b} \boldsymbol{E}_{\rm bb}(t-t') \boldsymbol{L}^{\rm bs} \\ & + \int_0^t ds \int_0^s dt' \Big[\big(\delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(s) \circ \delta \boldsymbol{x}^{\rm s\,T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t')\big) + \\ & \qquad \big(\delta \boldsymbol{x}^{\rm s\,T}(t') \boldsymbol{L}^{\rm sb} \boldsymbol{E}_{\rm bb}(s-t') \circ \delta \boldsymbol{x}^{\rm b\,T}(0) \boldsymbol{E}_{\rm bb}(s)\big) \Big] \boldsymbol{L}^{\rm bb,b} \boldsymbol{E}_{\rm bb}(t-s) \boldsymbol{L}^{\rm bs} \end{aligned} \tag{E.38}$$

Recalling equations (E.24), (E.27) and (E.29), one can write

$$m_4' = \boldsymbol{r}_{1\,\rm GVA}^T(t) \tag{E.39}$$

From (E.34) and (E.39) we have

$$m_1' + m_2' + m_3' = \boldsymbol{\mathcal{M}}_{\rm GVA}^T(t) + \boldsymbol{r}_{1\,\rm GVA}^T(t) \tag{E.40}$$

Let us now compare term by term the nonlinear memory from projection methods (E.10) and from the GVA (E.34): one has immediately $m_1 = m_1'$ (see (E.11) and (E.35)). We apply next the identity

$$(\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\rm ss} \boldsymbol{L}^{\rm ss,\cdot b} = \partial_{t'} (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} - (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,\cdot b} \tag{E.41}$$

to manipulate the sum of m_2 (E.12) and m_3 (E.13), as follows

$$\begin{aligned} m_2 + m_3 = {} & \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\rm ss} \boldsymbol{L}^{\rm ss,\cdot b} [\boldsymbol{E}_{\cdot\rm b,b}(t-t') \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t-t') \boldsymbol{L}^{\cdot\rm b,s}] = \\ = {} & \int_0^t dt' \big[\partial_{t'} (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} - (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,\cdot b}\big] \big[\boldsymbol{E}_{\cdot\rm b,b}(t-t') \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t-t') \boldsymbol{L}^{\cdot\rm b,s}\big] = \\ = {} & - \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} \partial_{t'} [\boldsymbol{E}_{\cdot\rm b,b}(t-t') \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t-t') \boldsymbol{L}^{\cdot\rm b,s}] \\ & + (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} [\boldsymbol{E}_{\cdot\rm b,b}(t-t') \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t-t') \boldsymbol{L}^{\cdot\rm b,s}] \Big|_0^t \\ & - \int_0^t dt' (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t'})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,\cdot b} [\boldsymbol{E}_{\cdot\rm b,b}(t-t') \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t-t') \boldsymbol{L}^{\cdot\rm b,s}] \end{aligned} \tag{E.42}$$

In the third line, an integration by parts has been implemented and $-\partial_{t'}$ can be further substituted by ∂_t as it acts on the difference $(t-t')$. Let us the consider the boundary term from this integration (fourth line)

$$\begin{aligned} (\delta \boldsymbol{x}^{\cdot\cdot\,T}(0) e^{\boldsymbol{L}^{\cdot,\cdot\cdot} t})_{\cdot\rm b} \boldsymbol{L}^{\cdot\rm b,s} - \delta \boldsymbol{x}^{\cdot\rm b\,T}(0) \big(\boldsymbol{E}_{\cdot\rm b,b}(t) \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t) \boldsymbol{L}^{\cdot\rm b,s}\big) \\ = m_3' - \delta \boldsymbol{x}^{\cdot\rm b\,T}(0) \big(\boldsymbol{E}_{\cdot\rm b,b}(t) \boldsymbol{L}^{\rm bs} + \boldsymbol{E}_{\cdot\rm b,\cdot b}(t) \boldsymbol{L}^{\cdot\rm b,s}\big) \end{aligned} \tag{E.43}$$

where we used (E.37), $\boldsymbol{E}_{\mathrm{sb,b}}(0) = \boldsymbol{E}_{\mathrm{bb,b}}(0) = \mathbb{0}$, $\boldsymbol{E}_{\mathrm{sb,sb}}(0) = \boldsymbol{E}_{\mathrm{bb,bb}}(0) = \mathbb{1}$ and the fact that all the terms in $e^{\boldsymbol{L}^{\cdot,\cdot\cdot}t'}$ at $t' = 0$ give $\mathbb{1}$. Equation (E.42) can be rewritten

$$\begin{aligned} m_2 + m_3 = & m_3' - \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b,s}}) \\ & + \int dt'(\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot,\cdot\cdot}t'})_{\cdot\mathrm{b}}\partial_t[\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t')\boldsymbol{L}^{\cdot\mathrm{b,s}}] \\ & - \int dt'(\delta\boldsymbol{x}^{\cdot\cdot\,T}(0)e^{\boldsymbol{L}^{\cdot,\cdot\cdot}t'})_{\cdot\mathrm{b}}\boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}[\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t')\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t')\boldsymbol{L}^{\cdot\mathrm{b,s}}] \end{aligned} \tag{E.44}$$

In the second line of (E.44), we can apply the identity

$$\partial_t\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t') = \boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}\boldsymbol{E}_{\cdot\mathrm{b,b}}(t-t') + \boldsymbol{L}^{\cdot\mathrm{b,b}}\boldsymbol{E}_{\mathrm{bb}}(t-t') \tag{E.45}$$

which can be deduced from the properties (E.7) and from

$$\partial_t\boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t') = \boldsymbol{L}^{\cdot\mathrm{b},\cdot\mathrm{b}}\boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t-t') \tag{E.46}$$

As a consequence, also by recalling (E.36), we obtain

$$m_2 + m_3 = m_3' - \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b,s}}) + m_2' \tag{E.47}$$

In light of (E.10) and (E.40) we thus have that, symbolically,

$$\boldsymbol{\mathcal{M}}^T_{\mathrm{GVA}}(t) + \boldsymbol{r}^T_{1\,\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}^T_{\mathrm{proj}}(t) + \delta\boldsymbol{x}^{\cdot\mathrm{b}\,T}(0)(\boldsymbol{E}_{\cdot\mathrm{b,b}}(t)\boldsymbol{L}^{\mathrm{bs}} + \boldsymbol{E}_{\cdot\mathrm{b},\cdot\mathrm{b}}(t)\boldsymbol{L}^{\cdot\mathrm{b,s}}) \tag{E.48}$$

Here the second term on the r.h.s. is exactly the nonlinear part of the projection random force (E.14), which we here denote $\boldsymbol{r}^T_{1\,\mathrm{proj}}(t)$. Finally one obtains the equivalence at $\mathcal{O}(\delta^2)$ we aimed to prove, i.e.

$$\boldsymbol{\mathcal{M}}_{\mathrm{GVA}}(t) + \boldsymbol{r}_{1\,\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}_{\mathrm{proj}}(t) + \boldsymbol{r}_{1\,\mathrm{proj}}(t) \tag{E.49}$$

The comparison becomes even simpler if the bulk is assumed initially at steady state, i.e. $\delta\boldsymbol{x}^{\mathrm{b}}(0) \equiv 0$, as all the random force terms then vanish and $\boldsymbol{\mathcal{M}}_{\mathrm{GVA}}(t) = \boldsymbol{\mathcal{M}}_{\mathrm{proj}}(t)$.