Mathematical and Computer Modelling of Dynamical Systems
Vol. 00, No. 00, Month 200x, 1–28

RESEARCH ARTICLE

Balanced model reduction of partially-observed Langevin equations: an averaging principle

Carsten Hartmann

Freie Universität Berlin, Institut für Mathematik
Arnimallee 6, 14195 Berlin, Germany
(*Received 00 Month 200x; final version received 00 Month 200x*)

We study balanced model reduction of partially-observed stochastic differential equations of Langevin type. Upon balancing, the Langevin equation turns into a singularly perturbed system of equations with slow and fast degrees of freedom. We prove that in the limit of vanishing small Hankel singular values (i.e., for infinite scale separation between fast and slow variables), its solution converges to the solution of a reduced-order Langevin equation. The approach is illustrated with several numerical examples, and we discuss the relation to model reduction of deterministic control systems having an underlying Hamiltonian structure.

Keywords: Stochastic Langevin equation, model order reduction, balanced truncation, averaging principle, optimal prediction

AMS Subject Classification: 60H10, 70L05, 93B11, 93C05, 93C70.

1. Introduction

We present a method for the model reduction of partially-observed linear stochastic differential equations of second-order Langevin type. Linear Langevin equations appear in a variety of physical contexts, e.g., molecular dynamics [1–3] or material science [4, 5]. Our approach is based on the method of balanced truncation that is a widely used model reduction technique for deterministic linear control systems [6, 7]. Other than common spatial decomposition methods such as the Proper Orthogonal Decomposition or the Principal Component Analysis that aim at projecting the dynamics onto certain "high-energy" modes, balanced model reduction consists in finding a coordinate (i.e., balancing) transformation such that modes which are least sensitive to the input variable (controllability) also give the least output (observability) and therefore can be neglected [8, 9].

In terms of the stochastic Langevin equation balancing means that we seek a representation in which variables that are most sensitive to the random excitations by the noise are also strongly coupled to the observed process and vice versa; the dominant variables will then carry most of the statistical weight. However the truncation step is a bit more involved than in the deterministic case, for weak excitability (controllability) of modes does not imply "smallness", the reason being that the noise process that drives the dynamics is unbounded. As a consequence even degrees of freedom that are only weakly excitable will almost surely become infinitely large, and therefore cannot simply be discarded. Instead, as we will argue, the weakly

E-mail: chartman@mi.fu-berlin.de

ISSN: 1387-3954 print/ISSN 1744-5051 online

DOI: 10.1080/1387395YYxxxxxxxx
http://www.informaworld.com

excitable and observable modes must be replaced by their quasi-stationary probability distribution, which then yields a dimension-reduced closed-form system of equations for the dominant variables. The thus reduced system is again of Langevin form and inherits many properties of the original Langevin equations.

The article is organized as follows: In Section 2, we give a tutorial introduction to the stochastic Langevin model. Section 3 explains the idea of balancing stochastic differential equations. Our main result, the averaging principle for the Langevin equation, is derived and proved in Section 4. Section 5 concludes with two numerical examples. The Appendix contains a brief derivation of the Langevin equation from a heat bath model and some technical results that are used in the article.

2. Stochastic Langevin equation

A prevalent model for the dynamics of n, say, one-dimensional particles in a heat bath under a linear force is the second-order Langevin equation [3]

$$M\ddot{Q}(t) + \gamma\dot{Q}(t) + KQ(t) = \xi(t)\,. \tag{2.1}$$

It describes the motion of a conservative particle that is subject to dissipation and fluctuations originating from the bath. Here $Q = (Q_1, \ldots, Q_n) \in \mathbf{R}^n$ is the vector of particle configurations, $M = \mathrm{diag}(m_1, \ldots, m_n)$ is the positive definite mass matrix with m_i being the mass of the i-th particle, and $\gamma, K \in \mathbf{R}^{n\times n}$ denote the symmetric positive definite friction and stiffness matrices. The driving force ξ on the right hand side of the equation is an n-dimensional uncorrelated stochastic process (white noise) with the covariance matrix

$$\mathbf{E}[\xi(t)\xi(t)^T] \propto \gamma$$

where, here and in the following, we use the symbol $\mathbf{E}[\cdot]$ to denote the expectation of a (measurable) stochastic process over all its possible realizations. The last equation is the celebrated Einstein-Smoluchowski or fluctuation-dissipation (FD) relation [2, 10, 11]; roughly speaking, it says that the excitations coming from the heat bath are balanced by the viscous friction, i.e., the energy dissipation into the bath which implies the existence of a stationary probability distribution (see Sec. 2.1 below). For general modelling purposes, of course, both friction and noise contributions may be considered independent.

The Langevin equation (2.1) often appears as the Markovian approximation of the so-called generalized Langevin equation [12]

$$M\ddot{Q}(t) + \int_0^t \hat{\gamma}(t-s)\dot{Q}(s)\,ds + KQ(t) = \zeta(t) \tag{2.2}$$

which includes memory effects that may be caused by slowly-decaying correlations between the heat bath and the particles. In this case, ζ is a "coloured" noise process with an autocovariance that is, again, determined by the FD relation,

$$\mathbf{E}[\zeta(t)\zeta(s)^T] \propto \hat{\gamma}(t-s)\,.$$

If the friction or memory kernel $\hat{\gamma}$ decays quickly, a good approximation is

$$\hat{\gamma}(t) = \gamma\delta(t)$$

which then, by the FD relation, implies (2.1); see also the appendix for a simple derivation of a simplified version of (2.1) from deterministic Newtonian dynamics.

Partially-observed Langevin equation In this article we will consider the linear Langevin equation (2.1) in a slightly more general form (still in its Markovian variant though). In particular we will allow the covariance matrix of the noise to be independent of the friction coefficient. To this end it is convenient to recast the second-order Langevin as a first order system: given a quadratic Hamiltonian

$$H\colon \mathbf{R}^{2n} \supseteq \mathbf{X} \to \mathbf{R}\,, \quad H(x) = \frac{1}{2} x \cdot Ex \tag{2.3}$$

with a matrix $E = E^T \succ 0$ ("$\succ$" means positive definite) and the notation $x \cdot y = x^T y$, we consider a stochastic Itô differential equation of the form

$$dX_t = (J - D)\,\nabla H(X_t)\,dt + \sqrt{\epsilon} B dW_t\,, \quad X_0 = x\,, \tag{2.4}$$

where W_t denotes standard Brownian motion in $\mathbf{R}^m$ (Wiener process), $J \in \mathbf{R}^{2n\times 2n}$ is the invertible skew-symmetric structure matrix, $D \in \mathbf{R}^{2n\times 2n}$ is the symmetric positive semidefinite friction matrix, and $B \in \mathbf{R}^{2n\times m}$ is the noise matrix. The gradient $\nabla = (\partial/\partial x_1, \ldots, \partial/\partial x_{2n})^T$ is understood as a column vector. Finally, $\epsilon > 0$ is a scalar parameter that controls the noise intensity (temperature).

We also call (2.4) a *Langevin equation*, and it can be readily seen that, for $m = n$ and a suitable choice of the matrices E, D, B, equation (2.4) reduces to (2.1). In deterministic control theory, equations of the form (2.4) are also known by name of *port-controlled Hamiltonian systems with dissipation.*

As is common in linear systems theory, we assume that not all states $x \in \mathbf{X}$ are relevant or accessible in any practical situation, so we augment our Langevin equation (2.4) by a linear output equation

$$Y_t = C\nabla H(X_t) \tag{2.5}$$

with $C \in \mathbf{R}^{l\times 2n}$ being the output matrix. (Note that $Y_t = CEX_t$.)

Moments of the Langevin process The Langevin process is Gaussian (as well as the observed process), i.e., it is completely determined by its mean and its covariance. Using Itô's formula and the shorthand $A = (J - D)E$, the solution of the Langevin equation (2.4) can be expressed by the stochastic integral

$$X_t = \exp(At)x + \sqrt{\epsilon} \int_0^t \exp(A(t-s))B\,dW_s$$

By the properties of the Itô integral [14, Theorem 3.2.1], the martingale term in the last equation, i.e., the stochastic integral has mean zero, so we find[1]

$$\mathbf{E}[X_t] = \exp(At)x\,. \tag{2.6}$$

[1]Throughout the article we will use the symbol $\mathbf{E}[\cdot]$ to denote the expectation of a stochastic process over all its possible realizations.

The covariance $\mathrm{cov}(X) = \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])^T]$ is given by

$$\begin{aligned}\mathrm{cov}(X_t) &= \epsilon \mathbf{E}\left[\int_0^t \exp(A(t-s))B\, dW_s \int_0^t dW_s^T B^T \exp(A^T(t-s))\right] \\ &= \epsilon \int_0^t \exp(A(t-s))BB^T \exp(A^T(t-s))\, ds \\ &= \epsilon \int_0^t \exp(As)BB^T \exp(A^T s)\, ds\end{aligned} \tag{2.7}$$

where we have used the (second) Itô isometry in the second line to replace the stochastic integral by a time integral ([14, Lemma 3.1.5]). If the matrix A is asymptotically stable in the sense of Lyapunov, i.e., all its eigenvalues have strictly negative real part (in brief: A is stable), then

$$\mathbf{E}[X_t] \to 0 \quad \text{and} \quad \mathrm{cov}(X_t) \to \epsilon \int_0^\infty \exp(As)BB^T \exp(A^T s)\, ds$$

as $t \to \infty$. Hence, using the linearity of Y_t and integrating by parts in the last equation, it follows that $\mathbf{E}[Y_t] \to 0$ and $\mathrm{cov}(Y_t) \to CEQEC^T$ as $t \to \infty$ with Q being the unique symmetric positive semidefinite solution of the Lyapunov equation

$$AQ + QA^T = -\epsilon BB^T\,.$$

2.1. *Fokker-Planck picture*

For our purposes it will be convenient to have an alternative, yet equivalent representation of the Langevin equation (2.4). To this end we introduce the infinitesimal generator of the process that is given by the second-order differential operator

$$L = \frac{\epsilon}{2}\mathrm{tr}\left(BB^T\nabla^2\right) + (J - D)\,\nabla H \cdot \nabla \tag{2.8}$$

where we use the notation $\nabla^2 f$ to denote the Hessian matrix of a function f. Throughout this article we will also use the convenient notation

$$BB^T : \nabla^2 = \mathrm{tr}\left(BB^T\nabla^2\right)$$

to denote the inner product between matrices (here: BB^T and ∇^2). We will also need the so-called Fokker-Planck or Kolmogorov forward operator

$$L^* = \frac{\epsilon}{2}BB^T : \nabla^2 - (J - D)\,\nabla H \cdot \nabla + D : \nabla^2 H \tag{2.9}$$

that is the formal adjoint of L with respect to the scalar product

$$(u, v) = \int_{\mathbf{X}} u(x)v(x)\, dx$$

between functions u and v; in other words, $(Lu, v) = (u, L^*v)$. Now, if X_t is the solution of (2.4) with sharp initial condition $X_0 = x$, then the distribution

$$\rho(z, t)dz = \mathbf{P}[X_t \in [z, z + dz)|X_0 = x]$$

of X_t will be governed by the Fokker-Planck equation [14]

$$\partial_t \rho(z,t) = L^* \rho(z,t)\,, \quad \lim_{t\to 0} \rho(z,t) = \delta(z-x)\,. \tag{2.10}$$

Exploiting the fact that X_t is a Gaussian process, and knowing its mean (2.6) and its covariance (2.7), we conclude that the solution to (2.10) reads

$$\rho(z,t) = \mathcal{N}\left(\exp(At)x,\, \epsilon \int_0^t \exp(As) BB^T \exp(A^T s)\, ds\right)$$

where $\mathcal{N}(m,\Sigma)$ denotes a normal distribution with mean m and covariance Σ.

Boltzmann-Gibbs distribution A typical problem in statistical mechanics consists in computing expectation values of certain observables with respect to the stationary solutions of the Fokker-Planck equation, i.e., solutions satisfying $L^*\rho = 0$. One such instance is the Boltzmann-Gibbs distribution $\rho_\infty \propto \exp(-H/\epsilon)$ that is a stationary solution of (2.10) if and only if the fluctuation-dissipation relation

$$2D = BB^T \tag{2.11}$$

holds. Indeed,

$$\begin{aligned} L^*\rho_\infty &= \left[\frac{1}{2} BB^T : \left(\frac{\nabla H \otimes \nabla H}{\epsilon} - \nabla^2 H\right) - \frac{\nabla H \cdot D \nabla H}{\epsilon} + D : \nabla^2 H\right] \rho_\infty \\ &= \left[\frac{\nabla H \cdot BB^T \nabla H}{2\epsilon} - \frac{1}{2} BB^T : \nabla^2 H - \frac{\nabla H \cdot D \nabla H}{\epsilon} + D : \nabla^2 H\right] \rho_\infty \end{aligned}$$

which, by positivity of $\rho_\infty \propto \exp(-H/\epsilon)$, entails $\partial_t \rho_\infty = 0$ if and only if $2D = BB^T$.

A much stronger requirement for the probability distribution ρ_∞ is ergodicity, which means that it is the *only* stationary solution of (2.10) that is approached for *any* initial distribution $\rho_0 = \rho(\cdot, 0)$ as time goes to infinity. In this case Birkhoff's ergodic theorem implies that time averages coincide with ensemble averages, i.e.,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t) dt = \int_{\mathbf{X}} f d\mu$$

for almost all initial values $X_0 = x$ where $f\colon \mathbf{X} \to \mathbf{R}$ is any integrable function for which the rightmost integral exists and we have defined

$$d\mu(x) = \rho_\infty(x)\, dx\,, \quad \int_{\mathbf{X}} \rho_\infty(x)\, dx = 1\,.$$

For our linear Langevin system (2.3)–(2.4) to be ergodic, it is sufficient that [15]

$$\operatorname{rank}\left(B|\, AB|\, A^2B|\, \ldots |\, A^{2n-1}B\right) = 2n\,,$$

i.e., the system must be completely controllable (provided that A is stable).

3. Balanced model reduction

Balanced model reduction for linear control systems goes back to [7] and is based on notions of controllability and observability. Roughly speaking, the idea is to reduce the system to those states only that share "good" controllability and observability properties. Given a stable linear system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t)\,, \quad x(0) = x \\ y(t) &= Cx(t) \end{aligned} \tag{3.1}$$

where, say, $x \in \mathbf{X} \subseteq \mathbf{R}^d$, $u \in \mathbf{R}^m$ and $y \in \mathbf{R}^l$, we have:

Definition 3.1: Following [16], controllability and observability functions associated with the system (3.1) are defined as

$$L_c(x) = \inf_{u \in L^2(-\infty,0]} \left\{ \int_{-\infty}^{0} |u(t)|^2\, dt \,:\, x = \int_{-\infty}^{0} \exp(A(t-s))Bu(s)\, ds \right\} \tag{3.2}$$

and

$$L_o(x) = \left\{ \int_{0}^{\infty} |y(t)|^2\, dt \,:\, y(t) = C\exp(At)x \right\} \tag{3.3}$$

The value of the controllability function L_c is the minimum control effort needed to steer the system from the zero state at $t = -\infty$ to a prescribed state $x \in \mathbf{X}$ at $t = 0$; note that L_c may be infinite if a state is uncontrollable. Conversely the observability function measures the output energy generated by a particular initial state $x \in \mathbf{X}$. The following standard result goes back to [6, 7].

Theorem 3.2 For a stable system, controllability and observability functions are given by

$$L_c(x) = x \cdot W_c^{-1} x\,, \quad L_o(x) = x \cdot W_o x$$

with the controllability Gramian W_c and the observability Gramian W_o being the unique and symmetric solutions of the Lyapunov equations

$$AW_c + W_c A^T = -BB^T\,, \quad A^T W_o + W_o A = -C^T C\,.$$

If the system is completely controllable and observable, i.e., $W_c, W_o \succ 0$, then there exists a coordinate (or *balancing*) transformation $x \mapsto Tx$ such that

$$T^{-1} W_c T^{-T} = T^T W_o T = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)\,.$$

The $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_d > 0$ are called Hankel singular values (HSV) and are independent of the choice of coordinates.

In the balanced representation $(A, B, C) \mapsto (T^{-1}AT, T^{-1}B, CT)$, states $x \in \mathbf{X}$ that are easily controllable have also good observability properties and vice versa. Accordingly *balanced truncation* consists in first balancing the system (3.1), and then projecting out the least observable and controllable states which have little effect on the input-output behaviour of the system [9].

3.1. *Balancing of the Langevin equation*

Before we discuss balanced truncation for the Langevin system (2.4)–(2.5) we have to make precise in which sense states are controllable or observable. In

$$\begin{aligned} dX_t &= (J-D)\,\nabla H(X_t)\,dt + \sqrt{\epsilon}B dW_t\,, \quad X_0 = x \\ Y_t &= C\nabla H(X_t) \end{aligned} \tag{3.4}$$

the driving process is incremental Brownian motion (i.e., white noise) rather than a smooth control variable. Clearly, Brownian motion is not differentiable, so we have to adapt the definition of the controllability function (3.2) suitably. To this end we replace the Brownian motion W_t by a polygonal approximation: let $\Delta^N = \{0 = t_0, t_1, t_2 \ldots, t_N = T\}$ be a partition of the interval $[0, T]$, and set

$$\hat{W}_t = W_{t_k} + \frac{t - t_k}{t_{k+1} - t_k}\left(W_{t_{k+1}} - W_{t_k}\right)\,, \quad t \in [t_k, t_{k+1}]\,.$$

We may now replace the Brownian increment dW_t in (3.4) by its approximants $d\hat{W}_t$ to obtain an auxiliary, Langevin-type system of the form

$$\begin{aligned} \dot{x}(t) &= (J-D)\,\nabla H(x(t)) + \sqrt{\epsilon}Bu(t)\,, \quad x(0) = x \\ y(t) &= C\nabla H(x(t)) \end{aligned} \tag{3.5}$$

where $u = d\hat{W}/dt$ denotes the piecewise constant time derivative of the polygonal approximant. One can show that the solutions of (3.5) converge almost surely to the solutions of the original Langevin equation as $N \to \infty$, i.e., as the mesh size of the partition Δ^N goes to zero [17]; for more details the interested reader is referred to the appendix (cf. also [18]).

Rate function The above considerations suggest that the reachability properties of the Langevin equation (3.4) may well be studied in terms of the associated control system (3.5). Roughly speaking we say that a state $x \in \mathbf{X}$ is less sensitive to the stochastic white noise than another state $x' \in \mathbf{X}$ if its controllability function is larger, i.e., if $L_c(x) > L_c(x')$. More precisely, we exploit the fact that the step functions are dense in the space of square integrable functions and state the following definition.

Definition 3.3: The rate function is defined as (see, e.g., [19])

$$L_r(x) = \inf_{u \in L^2[0,T]} \left\{ \int_0^T |u(t)|^2\,dt \,:\, x = \phi_0(t;u) \right\} \tag{3.6}$$

where $\phi_x(\cdot;u)\colon [0,T] \to \mathbf{X}$ with

$$\phi_x(t;u) = \exp(At)x + \sqrt{\epsilon}\int_0^t \exp(A(t-s))Bu(s)\,ds$$

denotes the solution of the deterministic system (3.5).

We declare that $L_r(x) = \infty$ when no admissible control $u \in L^2[0,T]$ exists. As

is proved in Appendix C the rate function assumes the familiar form, namely,

$$L_r(x) = x \cdot K_T^{-1} x\,,$$

with

$$K_T = \epsilon \int_0^T \exp(As) B B^T \exp(A^T s)\, ds\,.$$

being the covariance matrix of the process X_t at time $t = T$. The name *rate function* for L_r is owed to its use in Large Deviations Theory [20]; for our purposes it suffices to say that the rate function measures the minimum noise that is needed for the process to reach $x \in \mathbf{X}$ after time $t = T$, when it was started at $x = 0$ at time $t = 0$.

Balanced representation We make the following standing assumptions for (3.4):

(1) The matrix $A = (J - D)\nabla^2 H$ is stable, i.e., all eigenvalues of A are lying in the open left-half complex plane.
(2) The system is completely controllable and observable, i.e., both the controllability matrix $(B\ AB\ A^2B \dots A^{2n-1}B)$ and the observability matrix $(C\ CA\ CA^2 \dots CA^{2n-1})^T$ have maximum rank $2n$.

Now let W_c, W_o be the unique symmetric solutions of the Lyapunov equations

$$AW_c + W_c A^T = -\epsilon B B^T\,, \quad A^T W_o + W_o A = -E C^T C E\,. \tag{3.7}$$

Setting $\epsilon = 0$ in (3.4) it readily follows that W_o is the observability Gramian of our Langevin system. Since the matrix A is stable as we always assume, we moreover observe (upon integrating by parts) that the rate Gramian K_T of Lemma C.1 converges to the W_c as $T \to \infty$. By ergodicity the asymptotic rate Gramian (or controllability Gramian) $W_c = K_\infty$ is unique and symmetric positive definite. Also the observability Gramian W_o is positive definite if we assume that the Langevin equation is completely observable in which case a balancing transformation T is available that makes both W_c and W_o equal and diagonal, i.e.,

$$T^{-1} W_c T^{-T} = T^T W_o T = \mathrm{diag}(\sigma_1, \dots, \sigma_{2n})\,.$$

Now let $S = T^{-1}$ denote the inverse transformation. Using Itô's formula we easily see that in the balanced variables $z = Sx$ the Langevin system (3.4) reads[1]

$$\begin{aligned} dZ_t &= (\tilde{J} - \tilde{D})\nabla \tilde{H}(Z_t) dt + \sqrt{\epsilon}\tilde{B} dW_t \\ Y_t &= \tilde{C}\nabla \tilde{H}(Z_t) \end{aligned} \tag{3.8}$$

with the balanced Hamiltonian $\tilde{H}(z) = H(Tz)$, i.e.,

$$\tilde{H}(z) = \frac{1}{2} z \cdot \tilde{E} z\,, \quad \tilde{E} = T^T E T\,, \tag{3.9}$$

[1]In case of a linear transformation Itô's formula boils down to standard chain rule.

and the transformed coefficients

$$\tilde{J} = SJS^T\,, \quad \tilde{D} = SDS^T\,, \quad \tilde{B} = SB\,, \quad \tilde{C} = CS^T\,. \tag{3.10}$$

Hence the balancing transformation leaves the structure of the Langevin equation invariant. Since moreover $\tilde{A} = (\tilde{J} - \tilde{D})\tilde{E} = T^{-1}AT$ stability and complete controllability/observability are trivially preserved.

4. Balanced averaging of the Langevin equation

By linearity the observed process

$$Y_t = C\nabla H(X_t)\,, \quad X_t = \exp(At)x + \sqrt{\epsilon}\int_0^t \exp(A(t-s))B\,dW_s$$

is a Gaussian process and is invariant under coordinate transformations $x \mapsto Tx$, i.e., the original system (3.4) and the balanced system (3.8) generate identical observation sequences. In the balanced representation those states that are least sensitive to the driving white noise also generate the least output which is why we expect that truncating these states will have no major effect on the observed process. Accordingly the aim of this section is to derive a Langevin equation

$$\begin{aligned} dQ_t &= (\hat{J} - \hat{D})\nabla\hat{H}(Q_t)dt + \sqrt{\epsilon}\hat{B}dW_t \\ \hat{Y}_t &= \hat{C}\nabla\hat{H}(Q_t) \end{aligned}$$

on the state space $\mathbf{Q} \subseteq \mathbf{R}^d$, $q \in \mathbf{Q}$ that has much smaller dimension than $\mathbf{X} \subseteq \mathbf{R}^{2n}$ while $\hat{Y}_t \approx Y_t$ in a suitable stochastic sense (precise statements will be given in Theorem 4.1 below).

4.1. *Small parameters*

We shall now explain, starting from a balanced representation, how to systematically derive a reduced Langevin equation. Let W_c, W_o be the two symmetric and positive definite Gramians defined by (3.7). Since we assume that the Langevin equation is completely controllable (ergodic) and observable, we can employ a Cholesky factorization of the two Gramians $W_c, W_o \succ 0$,

$$W_c = XX^T\,, \quad W_o = YY^T\,,$$

and do a singular value decomposition (SVD) of the full-rank matrix Y^TX, i.e.,

$$Y^TX = U\Sigma V^T = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}. \tag{4.1}$$

The partitioning $\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$ and $\Sigma_2 = \mathrm{diag}(\sigma_{d+1}, \ldots, \sigma_{2n})$ indicates which singular values are important and which are negligible. The remaining matrices satisfy $U_1^TU_1 = V_1^TV_1 = \mathbf{1}_{d\times d}$ and $U_2^TU_2 = V_2^TV_2 = \mathbf{1}_{r\times r}$ with $r = 2n - d$. In terms of the SVD, the balancing transformation T and its inverse $S = T^{-1}$ read

$$T = XV\Sigma^{-1/2}\,, \quad S = \Sigma^{-1/2}U^TY^T \tag{4.2}$$

as can be readily verified. In the balanced coordinates $z = Sx$, the states that are least influenced by the input also have the least influence on the output as readily follows from Theorem 3.2.

Now suppose that $\sigma_{d+1} \ll \sigma_d$. As the Hankel singular values (HSV) are coordinate invariant, the $\sigma_{d+1}, \dots, \sigma_{2n} > 0$ may serve a dimensionless small parameters. Moreover equation (4.2) relates the balancing transform explicitly with the HSV which tells us where the small parameters enter the balanced equations. The idea is to study the limit of vanishing small singular values. As has been shown [21] for deterministic systems of the form (3.5) this gradually eliminates the least observable and controllable states, thereby confining the system to the dominant subspace of the largest HSV.

To understand the idea of the *confinement limit* $\Sigma_2 \to 0$, we suppose that d is even and scale the HSV uniformly according to

$$(\sigma_1, \dots, \sigma_d, \sigma_{d+1}, \dots, \sigma_{2n}) \mapsto (\sigma_1, \dots, \sigma_d, \lambda\sigma_{d+1}, \dots, \lambda\sigma_{2n}),$$

i.e., in (4.1)–(4.2) we replace Σ_2 by $\lambda\Sigma_2$ and study the limit $\lambda \to 0$. By (4.2) the thus scaled balancing matrices are readily shown to be

$$S(\lambda) = \begin{pmatrix} S_{11} & S_{12} \\ \lambda^{-1/2} S_{21} & \lambda^{-1/2} S_{22} \end{pmatrix}, \quad T(\lambda) = \begin{pmatrix} T_{11} & \lambda^{-1/2} T_{12} \\ T_{21} & \lambda^{-1/2} T_{22} \end{pmatrix}.$$

As we know from the previous section, balancing $x \mapsto S(\lambda)x$ implies that the coefficients in the Langevin equation (3.4) transform according to

$$J - D \mapsto S(\lambda)(J - D)S(\lambda)^T, \quad B \mapsto S(\lambda)B, \quad C \mapsto CS(\lambda)^T,$$

by which (3.8) turns into a system of the form

$$\begin{aligned} dZ_t^\lambda &= (\tilde{J}^\lambda - \tilde{D}^\lambda)\nabla\tilde{H}^\lambda(Z_t^\lambda)dt + \sqrt{\epsilon}\tilde{B}^\lambda dW_t \\ Y_t^\lambda &= \tilde{C}^\lambda \nabla\tilde{H}^\lambda(Z_t^\lambda). \end{aligned}$$

Here $\tilde{H}^\lambda(z) = H(T(\lambda)z)$ is the balanced Hamiltonian

$$\tilde{H}^\lambda(z) = \frac{1}{2} z \cdot \tilde{E}^\lambda z, \quad \tilde{E}^\lambda = \begin{pmatrix} \tilde{E}_{11} & \lambda^{-1/2}\tilde{E}_{12} \\ \lambda^{-1/2}\tilde{E}_{21} & \lambda^{-1}\tilde{E}_{22} \end{pmatrix},$$

and the remaining coefficients are given by

$$\tilde{J}^\lambda - \tilde{D}^\lambda = \begin{pmatrix} \tilde{J}_{11} - \tilde{D}_{11} & \lambda^{-1/2}(\tilde{J}_{12} - \tilde{D}_{12}) \\ \lambda^{-1/2}(\tilde{J}_{21} - \tilde{D}_{21}) & \lambda^{-1}(\tilde{J}_{22} - \tilde{D}_{22}) \end{pmatrix}$$

and

$$\tilde{B}^\lambda = \begin{pmatrix} \tilde{B}_1 \\ \lambda^{-1/2}\tilde{B}_2 \end{pmatrix}, \quad \tilde{C}^\lambda = \begin{pmatrix} \tilde{C}_1 & \lambda^{-1/2}\tilde{C}_2 \end{pmatrix}.$$

with $\tilde{J}$, $\tilde{D}$, $\tilde{B}$, $\tilde{C}$ denoting the unscaled coefficients as in (3.10), i.e., for $\lambda = 1$.

Upon setting $z = S(\lambda)x$ and introducing the shorthands

$$a_{ij}^{\lambda}(z_1, z_2) = (\tilde{J}_{ij} - \tilde{D}_{ij}) \left.\frac{\partial \tilde{H}^{\lambda}}{\partial z_j}\right|_{z_1,z_2}, \quad c_{ij}^{\lambda}(z_1, z_2) = \tilde{C}_i \left.\frac{\partial \tilde{H}^{\lambda}}{\partial z_j}\right|_{z_1,z_2}$$

the original Langevin equation (3.4) is therefore equivalent to the singularly perturbed system of equations

$$\begin{aligned} dZ_{1,t}^{\lambda} &= a_{11}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda})dt + \frac{1}{\sqrt{\lambda}} a_{12}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda})dt + \sqrt{\epsilon}\tilde{B}_1 dW_t \\ dZ_{2,t}^{\lambda} &= \frac{1}{\sqrt{\lambda}} a_{21}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda})dt + \frac{1}{\lambda} a_{22}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda})dt + \sqrt{\frac{\epsilon}{\lambda}}\tilde{B}_2 dW_t \\ Y_t^{\lambda} &= c_{11}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda}) + \frac{1}{\sqrt{\lambda}} c_{22}^{\lambda}(Z_{1,t}^{\lambda}, Z_{2,t}^{\lambda})\,. \end{aligned} \tag{4.3}$$

which will be our objects of interest.

Some preliminary considerations On any compact time interval $[0, T]$ and for $\lambda > 0$ finite, the process Z_t^{λ} has bounded first and second moments, so we conclude that $\mathrm{cov}(Z_{2,t}^{\lambda}) = \mathcal{O}(\lambda)$ as $\lambda \to 0$. This suggests to introduce new state variables

$$(q, p) = (z_1, \lambda^{-1/2} z_2)\,.$$

In terms of the scaled variables q, p our Langevin system (4.3) reads

$$\begin{aligned} dQ_t^{\lambda} &= a_{11}(Q_t^{\lambda}, P_t^{\lambda})dt + \frac{1}{\lambda} a_{12}(Q_t^{\lambda}, P_t^{\lambda})dt + \sqrt{\epsilon}\tilde{B}_1 dW_t \\ dP_t^{\lambda} &= \frac{1}{\lambda} a_{21}(Q_t^{\lambda}, P_t^{\lambda})dt + \frac{1}{\lambda^2} a_{22}(Q_t^{\lambda}, P_t^{\lambda})dt + \frac{\sqrt{\epsilon}}{\lambda}\tilde{B}_2 dW_t \\ Y_t^{\lambda} &= c_{11}(Q_t^{\lambda}, P_t^{\lambda}) + \frac{1}{\lambda} c_{22}(Q_t^{\lambda}, P_t^{\lambda}) \end{aligned} \tag{4.4}$$

with

$$a_{ij}(q, p) = (\tilde{J}_{ij} - \tilde{D}_{ij}) \left.\frac{\partial \tilde{H}}{\partial z_j}\right|_{q,p}, \quad c_{ij}(q, p) = \tilde{C}_i \left.\frac{\partial \tilde{H}}{\partial z_j}\right|_{q,p}$$

now being independent of λ. In the scaled variables (q, p) the Langevin equation (4.4) is more transparent than in terms of (z_1, z_2), for the coefficients a_{ij} and c_{ij} become independent of λ, so we readily see that (4.4) is an instance of a system with slow and fast degrees of freedom where q is slow and p is fast.

4.2. *The averaging principle*

Given (4.4), we seek an effective equation for the slow variables in the limit $\lambda \to 0$. Before we state our main result we shall try to built some intuition regarding this limit. First of all note that "dP_t/dt" is of the order $1/\lambda^2$, whereas Q_t has a slow drift of order 1, hence P_t appears as a fast random forcing to the slow process Q_t— in other words, the *time scale separation* between Q_t and P_t is of the order $1/\lambda^2$.

To make this precise we introduce a stretched time $\tau = t/\lambda^2$, so that, ignoring the auxiliary equation for the observable for the moment, (4.4) becomes[1]

$$\begin{aligned} dQ^\lambda_\tau &= \lambda^2 a_{11}(Q^\lambda_\tau, P^\lambda_\tau)d\tau + \lambda a_{12}(Q^\lambda_\tau, P^\lambda_\tau)d\tau + \sqrt{\epsilon}\lambda \tilde{B}_1 dW_\tau \\ dP^\lambda_\tau &= \lambda a_{21}(Q^\lambda_\tau, P^\lambda_\tau)d\tau + a_{22}(Q^\lambda_\tau, P^\lambda_\tau)d\tau + \sqrt{\epsilon}\tilde{B}_2 dW_\tau \,. \end{aligned}$$

Now letting $\lambda \to 0$, the last system of equations reduces to the *associated system*

$$\begin{aligned} dQ^\lambda_\tau &= 0 \\ dP^\lambda_\tau &= a_{22}(Q^\lambda_\tau, P^\lambda_\tau)d\tau + \sqrt{\epsilon}\tilde{B}_2 dW_\tau \end{aligned}$$

Hence, roughly speaking, equation (4.4) is of the form (in the limit $\lambda \to 0$)

$$\begin{aligned} dQ^\lambda_t &= a_{11}(Q^\lambda_t, P^q_{t/\lambda^2})dt + \frac{1}{\lambda} a_{12}(Q^\lambda_t, P^q_{t/\lambda^2})dt + \sqrt{\epsilon}\tilde{B}_1 dW_t \\ Y^\lambda_t &= c_{11}(Q^\lambda_t, P^q_{t/\lambda^2}) + \frac{1}{\lambda} c_{22}(Q^\lambda_t, P^q_{t/\lambda^2}) \end{aligned} \tag{4.5}$$

with the abbreviation $q = Q^\lambda_t$ and P^q_t as the solution of the associated (fast) system

$$dP^q_t = a_{22}(q, P^q_t)dt + \sqrt{\epsilon}\tilde{B}_2 dW_t \,.$$

Strictly speaking, we are dealing with a system exhibiting *three time scales*, namely, 1, $1/\lambda$ and $1/\lambda^2$. But as we will see in Section 4.3 below, the intermediate time scale $1/\lambda$ does not play a role, for the singular term proportional to $1/\lambda$ in (4.4) vanishes as $\lambda \to 0$. The associated system for fixed $Q^\lambda_\tau = q$ can be recast as

$$dP^q_t = (\tilde{J}_{22} - \tilde{D}_{22})\tilde{E}_{22}(P^q_t + \tilde{E}^{-1}_{22}\tilde{E}_{21}q)dt + \sqrt{\epsilon}\tilde{B}_2 dW_t \tag{4.6}$$

which implies that P^q_t is a Gaussian process with asymptotic mean

$$\lim_{t\to\infty} \mathbf{E}[P^q_t] = -\tilde{E}^{-1}_{22}\tilde{E}_{21}q \tag{4.7}$$

and covariance

$$\lim_{t\to\infty} \operatorname{cov}(P^q_t) = \epsilon \int_0^\infty \exp(\tilde{A}_{22}s)\tilde{B}_2\tilde{B}_2^T \exp(\tilde{A}^T_{22}s)\, ds \tag{4.8}$$

where, for the latter, we assume that $\tilde{A}_{22} = (\tilde{J}_{22} - \tilde{D}_{22})\tilde{E}_{22}$ is stable; cf. [21].

The idea of the averaging principle is to interpret the phase space of the fast dynamics as a fibre over the phase space of the slow ones. Then, rather than treating the fast motion explicitly, only its average influence on the slow dynamics is considered. But since the fast variables relax almost instantaneously to their stationary distribution (4.7)–(4.8), we might well replace them by their average values. In other words, if P^q_t is ergodic, taking the limit $\lambda \to 0$ in the auxiliary system (4.5) essentially amounts to replacing P^q_t by its asymptotic mean. Since

$$a_{12}(q, -\tilde{E}^{-1}_{22}\tilde{E}_{21}q) = 0\,, \quad c_{22}(q, -\tilde{E}^{-1}_{22}\tilde{E}_{21}q) = 0$$

[1]Formally, $dW_t \sim \sqrt{dt}$, hence W_t behaves like $W_t \mapsto \lambda W_{t/\lambda^2}$ under scaling $t \mapsto t/\lambda^2$.

the singular terms in (4.5) vanish and we end up with the averaged equation

$$dQ_t = (\tilde{J}_{11} - \tilde{D}_{11})(\tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21})Q_t dt + \sqrt{\epsilon}\tilde{B}_1 dW_t$$
$$Y_t = \tilde{C}_1(\tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21})Q_t\,.$$

As can be readily seen, the latter is of Langevin type with an effective Hamiltonian

$$\bar{H}(q) = \frac{1}{2}q \cdot \bar{E}_1 q\,, \quad \bar{E}_1 = \tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21}\,.$$

Notice that $\tilde{J}_{11} = -\tilde{J}_{11}^T$ and $\tilde{D}_{11} = \tilde{D}_{11}^T \succcurlyeq 0$ are simply the original structure and friction matrices restricted to the subspace of the most controllable and observable states. That is, in the limit of vanishing small HSV the dynamics collapse (in distribution) to the controllable and observable subspace. Moreover $E = E^T \succ 0$ implies $\bar{E}_1 = \bar{E}_1^T \succ 0$ for the Schur complement. As for the stability of the limiting system and the comparison with the deterministic system we refer to [21]. Now comes our main result that is proved in Section 4.3 below:

Theorem 4.1 Let Y_t^λ be the observed solution of the Langevin equation (4.4) with (non-explosive) initial conditions independent of λ. Then, as λ goes to zero, Y_t^λ converges in expectation to Y_t, i.e.,

$$\lim_{\lambda\to 0} \mathbf{E}\left[\sup_{t\in[0,T]}\left|Y_t^\lambda - Y_t\right|\right] = 0 \quad \forall T > 0\,.$$

where Y_t is a Markov process that is governed by the reduced Langevin equation

$$\begin{aligned} dQ_t &= (\tilde{J}_{11} - \tilde{D}_{11})\nabla\bar{H}(Q_t)dt + \sqrt{\epsilon}\tilde{B}_1 dW_t \\ Y_t &= \tilde{C}_1\nabla\bar{H}(Q_t) \end{aligned} \tag{4.9}$$

with the effective Hamiltonian

$$\bar{H}(q) = \frac{1}{2}q \cdot \bar{E}_1 q\,, \quad \bar{E}_1 = \tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21}\,. \tag{4.10}$$

The next statement is a straight consequence.

Corollary 4.1 In (3.4), let friction and noise coefficients satisfy the fluctuation-dissipation relation $2D = BB^T$. Then also $2\tilde{D}_{11} = \tilde{B}_1\tilde{B}_1^T$, and the reduced system admits an invariant measure that is given by the marginal Boltzmann measure

$$d\bar{\mu}(q) = \frac{1}{\bar{Z}}\exp(-\beta\bar{H}(q))\,dq\,, \quad \bar{Z} = \int_{\mathbf{R}^d}\exp(-\beta\bar{H}(q))\,dq\,.$$

Moreover $\bar{H}$ can be expressed as the thermodynamic free energy

$$\bar{H}(q) = -\epsilon\ln\mathbf{P}_\mu(q)\,, \quad \mathbf{P}_\mu(q) = \int_{\mathbf{X}}\delta(z_1 - q)d\mu$$

which is independent of ϵ.

Proof: The fluctuation-dissipation relation is a straight consequence of the definition of the balanced coefficients (3.10). The marginal property follows from the

complete controllability of the original system and completing the square in

$$\begin{aligned}\bar{H}(q) &= -\epsilon \ln \int_{\mathbf{X}} \delta(z_1 - q) d\mu(z_1, z_2) \\ &= -\epsilon \ln \frac{1}{Z} \int_{\mathbf{R}^{2n-d}} \exp(-\tilde{H}(q, z_2)/\epsilon) \sqrt{\det T}\, dz_2 \\ &= \frac{1}{2} q \cdot (\tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21}) q + \text{const.}\end{aligned}$$

with T in the second equation denoting the balancing transformation (4.2). □

4.3. *Proof of the averaging principle*

The proof relies on a systematic perturbation expansion of the Kolmogorov backward equation associated with (4.4) and follows essentially the lines of [30, 31]; see also [33] for an introduction to the general methodology.

Derivation of the limit equation We start with the formal justification of the asymptotic result (4.9) by doing a perturbative expansion of the Langevin equation in powers of the small parameter λ. To this end we observe that the infinitesimal generator (2.8) associated with the Langevin equation (4.4) splits according to

$$L^\lambda = L_0 + \frac{1}{\lambda} L_1 + \frac{1}{\lambda^2} L_2$$

with

$$\begin{aligned}L_0 &= \frac{\epsilon}{2} \tilde{B}_1 \tilde{B}_1^T : \frac{\partial^2}{\partial q^2} + a_{11}(q,p) \cdot \frac{\partial}{\partial q} \\ L_1 &= \epsilon \tilde{B}_1 \tilde{B}_2^T : \frac{\partial}{\partial p} \frac{\partial}{\partial q} + a_{12}(q,p) \cdot \frac{\partial}{\partial q} + a_{21}(q,p) \cdot \frac{\partial}{\partial p} \\ L_2 &= \frac{\epsilon}{2} \tilde{B}_2 \tilde{B}_2^T : \frac{\partial^2}{\partial p^2} + a_{22}(q,p) \cdot \frac{\partial}{\partial p}\end{aligned}$$

and the shorthands

$$\frac{\partial}{\partial q} = \left(\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_d} \right)^T, \quad \frac{\partial}{\partial p} = \left(\frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_{2n-d}} \right)^T .$$

Suppose L^λ is equipped with appropriate boundary conditions, and consider the following Cauchy problem (Kolmogorov backward equation)

$$\partial_t v^\lambda(q,p,t) = L^\lambda v^\lambda(q,p,t)\,, \quad v^\lambda(q,p,0) = f(q) \tag{4.11}$$

that is the adjoint of the Fokker-Planck equation (2.10) and that is fully equivalent to the Langevin equation (4.4). For v^λ we seek a perturbative expansion of the form

$$v^\lambda = v_0 + \lambda v_1 + \lambda^2 v_2 + \ldots .$$

Plugging the ansatz in the backward equation (4.11) and equating equal powers of λ yields a hierarchy of equations the first three of which are

$$L_2 v_0 = 0 \tag{4.12}$$

$$L_2 v_1 = -L_1 v_0 \tag{4.13}$$

$$L_2 v_2 = -L_0 v_0 - L_1 v_1 + \partial_t v_0 \,. \tag{4.14}$$

We proceed step by step: First of all, note that L_2 is a differential operator in p only and that, by stability of $(J-D)E$ in the Langevin equation, the null space of $\tilde{J}_{22} - \tilde{D}_{22}$ is empty (see [21]). By the assumption of complete controllability of (4.4), it follows that also the operator L_2 which is the infinitesimal generator of the fast process P_t^q satisfies a controllability condition. Following [30], the latter implies that the fast dynamics relax exponentially fast to their unique invariant probability distribution

$$\eta(\cdot; q) = \mathcal{N}\left(-\tilde{E}_{22}^{-1}\tilde{E}_{21} q,\, \epsilon \int_0^\infty \exp(\tilde{A}_{22} s)\tilde{B}_2 \tilde{B}_2^T \exp(\tilde{A}_{22}^T s)\, ds \right).$$

Hence, up to a normalization factor, the equation $L_2^* \rho = 0$ has the unique solution $\rho = \eta$ and therefore the only functions that solve (4.12) are independent of p and hence are of the form $v_0 = v_0(q,t)$.

Regarding the second equation, (4.13), the Fredholm alternative states [32] that

$$\operatorname{ran} L_2 = (\ker L_2^*)^\perp$$

where orthogonality is meant with respect to the L^2 scalar product. As a consequence of the Fredholm alternative, (4.13) has a solution if and only if the right hand side is orthogonal to the null space of L_2^*. But this is to say that the right hand side of (4.13) is zero when we integrate it against $d\nu_q(p) = \eta(p;q)dp$. As v_0 is independent of p we immediately see that

$$\int_{\mathbf{R}^{2n-d}} L_1 v_0 \, d\nu_q = 0 \,,$$

i.e., the solvability condition $L_1 v_0 \perp \ker L_2^*$ is met. To solve equation (4.13) for the unknown v_1 we follow [33] and observe that v_1 must be of the form

$$v_1(q,p,t) = \phi(q,p)^T \nabla v_0(q,t) + \psi(q,t)$$

where $\psi \in \ker L_2$ plays no role in what follows so we set it to zero. Equation (4.13) can now be recast as an equation for $\phi\colon \mathbf{X} \to \mathbf{R}^d$, the so-called cell problem

$$L_2 \phi = -a_{12}^T \,. \tag{4.15}$$

In (4.11), the initial condition is independent of λ, therefore $v_1(q,p,0) = 0$ which leaves the only possible choices $v_0 = c$ or $\phi = 0$. If we exclude the trivial stationary solution v_0 being constant, consistency of (4.15) requires that $a_{12} = 0$, i.e., the initial conditions for p are drawn from the equilibrium distribution ν_q of the fast process; see also the remark below. To conclude, the Fredholm alternative for

equation (4.14) entails the solvability condition

$$\int_{\mathbf{R}^{2n-d}} (\partial_t v_0 - L_0 v_0 - L_1 \phi \nabla v_0)\, d\nu_q = 0$$

which, for $\phi = 0$, can be recast as an equation for q, namely,

$$\partial_t v_0(q,t) = \left(\frac{1}{2} \tilde{B}_1 \tilde{B}_1^T : \nabla^2 + \bar{a}(q) \cdot \nabla \right) v_0(q,t)\,. \tag{4.16}$$

Here we have introduced the abbreviation $\bar{a}(q) = \bar{A}q$ with

$$\bar{A} = (\tilde{J}_{11} - \tilde{D}_{11})(\tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21})\,.$$

As can be readily checked, (4.9) is the stochastic Langevin equation associated with (4.16) which concludes the first part of the proof. □

Convergence issues Since Y_t^λ in the Langevin equation (4.4) is linear in both Q_t^λ and P_t^λ where the fast process P_t^λ relaxes exponentially fast to its stationary distribution as $\lambda \to 0$, it suffices to confine our attention to convergence of the slow process Q_t^λ. To this end consider the Poisson equation

$$\begin{aligned} L_2 \phi(q,p) &= a_{11}(q,p) + \frac{1}{\lambda} a_{12}(q,p) - \bar{a}(q) \\ \phi(\cdot,p) &\to 0 \text{ as } |p| \to \infty\,. \end{aligned}$$

By construction the right hand side of the Poisson equation averages to zero if we integrate it against the equilibrium distribution $d\nu_q$ of the fast dynamics. Since moreover the fast dynamics are ergodic, the coefficients of L_2 satisfy the usual controllability condition, and it has been shown in [31] that the equation is well-posed and has a smooth solution. Itô's formula then yields

$$d\phi_t^\lambda = L^\lambda \phi(Q_t^\lambda, P_t^\lambda)dt + \sqrt{\epsilon} \frac{\partial \phi}{\partial q}(Q_t^\lambda, P_t^\lambda) dB_t^q + \frac{\sqrt{\epsilon}}{\lambda} \frac{\partial \phi}{\partial p}(Q_t^\lambda, P_t^\lambda) dB_t^p\,.$$

where we have introduced the shorthands $dB_t^q = \tilde{B}_1 dW_t$ and $dB_t^p = \tilde{B}_2 dW_t$. Employing the Poisson equation, the first line of equation (4.4) can be recast as

$$dQ_t^\lambda = \left(L_2 \phi(Q_t^\lambda, P_t^\lambda) + \bar{a}(Q_t^\lambda) \right) dt + \sqrt{\epsilon}\, dB_t^q$$

which, together with $L^\lambda = L_0 + \lambda^{-1} L_1 + \lambda^{-2} L_2$ and the equation for ϕ_t^λ turns into

$$\begin{aligned} dQ_t^\lambda &= \left(\bar{a}(Q_t^\lambda) - (\lambda^2 L_0 + \lambda L_1)\phi(Q_t^\lambda, P_t^\lambda) \right) dt + \lambda^2 d\phi_t^\lambda \\ &\quad + \sqrt{\epsilon} \left(1 - \lambda^2 \frac{\partial \phi}{\partial q}(Q_t^\lambda, P_t^\lambda) \right) dB_t^q - \sqrt{\epsilon} \lambda \frac{\partial \phi}{\partial p}(Q_t^\lambda, P_t^\lambda) dB_t^p \\ &= \bar{a}(Q_t^\lambda)dt + \sqrt{\epsilon}\, dB_t^q + \lambda \left(dS_t^\lambda + \sqrt{\epsilon} dM_t^\lambda \right)\,. \end{aligned}$$

Here we have introduced the abbreviations

$$dS_t^\lambda = -\left(L_1 + \lambda L_0\right)\phi(Q_t^\lambda, P_t^\lambda)dt + \lambda\, d\phi_t^\lambda$$

and

$$dM_t^\lambda = -\lambda\frac{\partial\phi}{\partial q}(Q_t^\lambda, P_t^\lambda)dB_t^q - \frac{\partial\phi}{\partial p}(Q_t^\lambda, P_t^\lambda)dB_t^p\,.$$

Since ϕ is smooth and the solution $(Q_t^\lambda, P_t^\lambda)$ of equation (4.4) exists on any finite time interval and for all $\lambda > 0$, the process $\xi_t^\lambda = S_t^\lambda - S_0^\lambda$ given by

$$S_t^\lambda - S_0^\lambda = \int_0^t dS_\tau^\lambda$$

satisfies, with probability one and for all $\lambda > 0$,

$$\sup_{0\le t\le T}|\xi_t^\lambda| \le C_1\,. \tag{4.17}$$

By the second Itô isometry [14, Lemma 3.1.5], the Martingale term

$$M_t^\lambda = -\lambda\int_0^t \frac{\partial\phi}{\partial q}(Q_\tau^\lambda, P_\tau^\lambda)dB_\tau^q - \int_0^t \frac{\partial\phi}{\partial p}(Q_\tau^\lambda, P_\tau^\lambda)dB_\tau^p$$

has bounded quadratic variation, i.e., for all $\lambda > 0$ and $t < \infty$,

$$\mathbf{E}[|M_t^\lambda|^2] = \int_0^t \mathbf{E}\left[\lambda\left|\frac{\partial\phi}{\partial q}\Big|_s dB_s^q\right|^2 + 2\left(\frac{\partial\phi}{\partial q}\Big|_s dB_s^q\right)\cdot\left(\frac{\partial\phi}{\partial p}\Big|_s dB_s^p\right) + \left|\frac{\partial\phi}{\partial p}\Big|_s dB_s^p\right|^2\right] ds$$

is finite which implies that

$$\mathbf{E}[|M_t^\lambda|] \le C_2\,. \tag{4.18}$$

Setting $\zeta_t^\lambda = Q_t^\lambda - Q_t$ with Q_t denoting the limit process (4.9), we obtain

$$\zeta_t^\lambda = \int_0^t \bar{a}(Q_s^\lambda - Q_\tau)ds + \lambda S_t^\lambda + \lambda M_t^\epsilon\,.$$

Using the linearity of $\bar{a}$ and (4.17)–(4.18) the last equation yields

$$|\zeta_t^\lambda| \le C_3\int_0^t |\zeta_s^\lambda|ds + \lambda C_4$$

with constants C_3 and $C_4 = C_1 + C_2$ on any compact time interval $[0, T]$. Therefore

$$\begin{aligned}\mathbf{E}\left[\sup_{0\le t\le T}|\zeta_t^\lambda|\right] &\le C_5\left(\lambda + C_3\int_0^T \mathbf{E}[|\zeta_s^\lambda|]ds\right)\\ &\le C_6\left(\lambda + \int_0^T \mathbf{E}\left[\sup_{0\le t\le T}|\zeta_s^\lambda|\right]ds\right)\,,\end{aligned}$$

and it follows from the integral form of Gronwall's lemma [34, Appendix 5] that

$$\mathbf{E}\left[\sup_{0\le t\le T}|\zeta_t^\lambda|\right] \le \lambda \exp(C_7 T)\,.$$

Finally, noting that

$$Y_t^\lambda - Y_t = c_{11}(Q_t^\lambda, P_t^\lambda) + \frac{1}{\lambda}c_{22}(Q_t^\lambda, P_t^\lambda) - \tilde{C}_1\nabla\bar{H}(Q_t)$$

and iterating the Poisson equation argument with

$$\begin{aligned} L_2\psi(q,p) &= c_{11}(q,p) + \frac{1}{\lambda}c_{22}(q,p) - \tilde{C}_1\nabla\bar{H}(q) \\ \psi(\cdot,p) &\to 0 \text{ as } |p|\to\infty\,. \end{aligned}$$

yields Theorem 4.1. □

4.4. *Optimal prediction*

The averaging principle, Theorem 4.1, admits a nice variational interpretation within the framework of optimal prediction that is due to Chorin and co-workers [35]: Suppose we want to solve (3.4), but we do not know the initial value exactly. All we know is that the initial values follow some joint probability distribution ρ which we assume to be the unique invariant distribution

$$\rho = \mathcal{N}\left(0, \epsilon\int_0^\infty \exp(As)BB^T\exp(A^Ts)\,ds\right).$$

Further assume that we have identified the most controllable and observable variables by balancing the controllability (rate) and observability Gramians of (3.4). We shall call these variables the *resolved variables*; the remaining ones are called *unresolved variables*. By nature of the resolved variables, namely, being easily controllable and observable it is plausible that we can acquire knowledge about their initial values. Given $q = z_1 \in \mathbf{R}^d$, the distribution of the unresolved variables $z_2 \in \mathbf{R}^{2n-d}$ is given by the joint probability density ρ conditioned by z_1, i.e., $z_2 \sim \eta(\cdot;q)$ with

$$\eta(\cdot;q) = \mathcal{N}\left(-\tilde{E}_{22}^{-1}\tilde{E}_{21}q,\ \epsilon\int_0^\infty \exp(\tilde{A}_{22}s)\tilde{B}_2\tilde{B}_2^T\exp(\tilde{A}_{22}^Ts)\,ds\right).$$

The aim now is to obtain an optimal prediction of the observed process Y_t at time t in terms of the resolved variables Q_t given that we know $Q_0 = q$ at time $t = 0$. Clearly $Y_t = C\nabla H(X_t)$ depends upon the unresolved variables via the Langevin process X_t, so we seek an appropriate closure scheme. One way to close the equations consists in replacing the right hand side of (3.4) by its best-approximation as a function of the resolved variables. To this end, we define the conditional expectation

$$\mathbf{E}\left[f|q\right] = \int_{\mathbf{R}^{2n-d}} f\,d\nu_q\,, \quad d\nu_q(z_2) = \eta(z_2;q)dz_2\,.$$

It is easy to check that the conditional expectation is an orthogonal projection in the weighted Hilbert space L^2_ρ equipped with the inner product

$$(f,g)_\rho = \int_{\mathbf{R}^{2n-d}} fg\, d\rho\,.$$

That is, defining $(\Pi f)(q) = \mathbf{E}\,[f|q]$ we have $\Pi^2 = \Pi$ and $(\Pi f, g)_\rho = (f, \Pi g)_\rho$. By being an orthogonal projection, the conditional expectation satisfies the best-approximation property

$$\|f - \Pi f\|^2_\rho \leq \|f - g\|^2_\rho \quad \forall g = g(q)\,. \tag{4.19}$$

The norm $\|\cdot\|_\rho$ is the norm induced in L^2_ρ by the inner product $(\cdot,\cdot)_\rho$ and, by stationarity of ρ, it appears to be the most natural one for our purposes.

Now consider our Langevin equation (3.4); if, for all $t \geq 0$, the process X_t is distributed according to ρ, then its increments will be Gaussian distributed according to what stands right of the equality in (3.4). Thus the best-approximation as a function of the most controllable and observable degrees of freedom $z_1 = q$ (with $(z_1, z_2)^T = Tx$ and T being the balancing transformation) is obtained upon taking the conditional expectation $\Pi = \mathbf{E}\,[\cdot|q]$. It should not come as a surprise that this yields the following projected Langevin equation

$$\begin{aligned} dQ_t &= (\tilde{J}_{11} - \tilde{D}_{11})\nabla\bar{H}(Q_t)dt + \sqrt{\epsilon}\tilde{B}_1 dW_t \\ Y_t &= \tilde{C}_1\nabla\bar{H}(Q_t) \end{aligned}$$

with the effective Hamiltonian

$$\bar{H}(q) = \frac{1}{2} q\cdot\bar{E}_1 q\,, \quad \bar{E}_1 = \tilde{E}_{11} - \tilde{E}_{12}\tilde{E}_{22}^{-1}\tilde{E}_{21}\,.$$

Hence the limit equation in Theorem 4.1 is the best-approximation with respect to the norm $\|\cdot\|_\rho$ as a function of the most controllable and observable modes. Before we conclude this section a final remark is in order.

Remark Interestingly enough the limit of vanishing small Hankel singular values of a deterministic Langevin-like (or dissipative Hamiltonian) system of the form (3.5) that has been studied in [21] yields formally the same limit system, namely,

$$\begin{aligned} \dot{q}(t) &= (\tilde{J}_{11} - \tilde{D}_{11})\nabla\bar{H}(q(t)) + \sqrt{\epsilon}\tilde{B}_1 u(t) \\ y(t) &= \tilde{C}_1\nabla\bar{H}(q(t)) \end{aligned}$$

with the same effective Hamiltonian (4.10). The convergence $y^\lambda(t) \to y(t)$, however, is rather different as in the limit the negligible modes get "slaved" by the dominant ones and the system fully collapses to the controllable/observable subspace. No other dynamics remain. As has been shown, the reduced system nicely preserves stability and passivity and admits the usual H^∞ error bound for the corresponding transfer function.

5. Numerical illustration

It readily follows from the convergence proof of the averaging principle that the deviations between the original observed process and the reduced one are, at least formally, of order λ. Consequently we expect an error that is of the order of the negligible Hankel singular values, bearing resemblance to the typical Hankel norm bounds for transfer functions [9]. Note, however, that the scaling parameter λ is a fake parameter that was introduced in order to highlight the scaling properties of the equations; the real small parameters are the negligible σ_i.

In this section, we discuss two numerical examples the first of which is taken from the SLICOT library of benchmark examples [36]. The second is purely pedagogical, but it allows us to control the small parameter which in the example below is the vanishing mass of a particle that is subject to random forcing and friction.

5.1. *An example from structural mechanics*

The following example is an adaptation of [36]. Consider a second-order Langevin equation of the form

$$\begin{aligned} dQ_t &= V_t dt \\ M dV_t &= -\left(KQ_t + RV_t\right) dt + S dW_t \\ Y_t &= C_1 Q_t + C_2 V_t \end{aligned} \tag{5.1}$$

with symmetric and positive definite coefficients $M, R, K \in \mathbf{R}^{135\times 135}$ and matrices $S \in \mathbf{R}^{135\times 3}$ and $C_1, C_2 \in \mathbf{R}^{3\times 135}$. The equations are a model for structural vibrations of a mechanical device that is embedded in a thermal bath. In our particular case the device is an elastic beam that is part of the international space station (ISS). Equivalently the equations can be thought of as the Markovian limit of the generalized Langevin equation for the beam where the degrees of freedom of the surrounding material are implicitly described by the noise and the dissipation [37]. However no assumptions regarding fluctuation-dissipation relation are made.

Upon introducing the Hamiltonian

$$H\colon \mathbf{R}^{135} \times \mathbf{R}^{135} \to \mathbf{R}\,, \quad H(x_1, x_2) = \frac{1}{2} x_2 \cdot M^{-1} x_2 + \frac{1}{2} x_1 \cdot K x_1$$

the above system can be seen to be a Langevin equation of the form (3.4). Rate and observability Gramians are then computed as solutions to the Lyapunov equations

$$AK + KA^T = -BB^T\,, \quad A^T W_o + W_o A = -N^T N$$

with the matrices

$$A = \begin{pmatrix} \mathbf{0} & M^{-1} \\ -K & -RM^{-1} \end{pmatrix}, \quad B = \begin{pmatrix} \mathbf{0} \\ S \end{pmatrix}, \quad N = \begin{pmatrix} C_1 \; C_2 M^{-1} \end{pmatrix}.$$

The resulting Hankel singular values (HSV) are shown in Figure 1. The value of the 40-th singular value is below $5 \cdot 10^{-5}$ which is less than one thousand of the first one. Interestingly enough the dominant HSV decay in pairs which indicates that the skew-symmetric Hamiltonian part contributes most to this part of the dynamics.

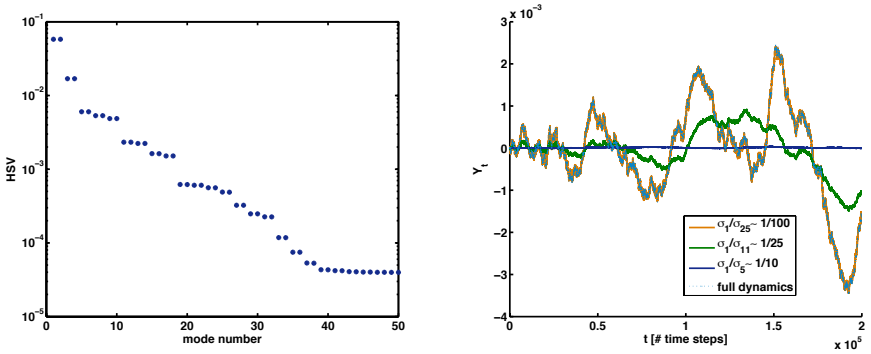

Figure 1. Left panel: The first 50 Hankel singular values of the ISS model (logarithmic scale). Right panel: Comparison between typical realizations of the third component of Y_t of the full system (dashed black curve) and reduced systems of degrees $d \in \{4, 10, 24\}$; for $d = 24$ the two full dynamics and the approximant (orange curve) are virtually indistinguishable.

Reduced dynamics We compare two types of realizations of the Langevin equation: those starting from the initial value $x = 0$ and those starting from non-zero initial conditions that are randomly drawn from a Gaussian distribution. The respective Langevin equations are discretized employing the Euler-Maruyama scheme with stable step size $h = 5 \cdot 10^{-6}$ (this is a rather conservative choice).

In case of zero initial conditions, the solution of (5.1) is

$$Y_t = N \int_0^t \exp(A(t-s))B \, dW_s \, .$$

Since $\mathbf{E}[Y_t] = 0$ the solution consists in Brownian fluctuations. The right panel of Figure 1 shows a comparison between a typical realization of the full, 270-dimensional dynamics and reduced systems of different degrees. For $d = 24.$, i.e., a ratio of $1/100$ between the first and the first neglected HSV the observable y_3 is almost exactly approximated by the averaged dynamics; for the first two components y_1, y_2 an even smaller degree with HSV ratio of about $1/25$ is sufficient ($d = 10$), so y_3 represents the worst-case scenario for our test system.

As for non-zero initial conditions and noise and friction coefficients as given, the solutions are dominated by the oscillatory Hamiltonian part. We distinguish two scenarios: In the first case, we randomly pick the value of the dominant variable and then compute the corresponding equilibrium value of the unresolved (i.e., negligible) variables; here the word *equilibrium* refers to the invariant distribution of the fast variables (recall the consideration from the beginning of Section 4.2). This scenario is contrasted with the case of completely random Gaussian initial values with zero mean (this is the "out-of-equilibrium" situation).

The left panel of Figure 2 shows a typical realization of the full dynamics and the approximant of degree $d = 10$. As one should expect the approximation is much better than for the fluctuations (i.e., for the second moment); the approximation is virtually indistinguishable from the original.

The approximation becomes slightly worse, however, if we drop the restriction on the equilibrium initial conditions as the right panel of Figure 2 shows for an approximant of degree $d = 24$. In this case a HSV ratio of about $1/1000 \ldots 1/2000$ is required before the relaxation of the fast dynamics to their invariant distribution kicks in, and the approximation becomes as good as for the equilibrium situation. This behaviour can easily be explained by bearing in mind that relaxation time of the fast dynamics (i.e., the time that is needed to equilibrate) is of the order of the first neglected HSV squared.

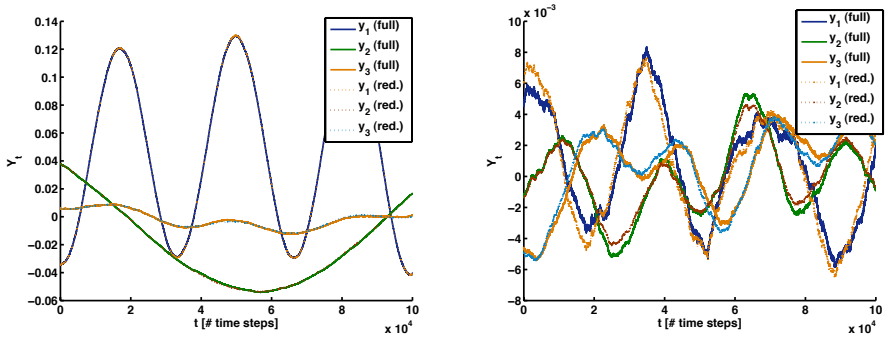

Figure 2. Left panel: Output Y_t of the full, 270-dimensional system and its 10-dimensional approximant (dashed curves) for random initial conditions restricted to the invariant subspace. Right panel: The same for Gaussian random initial conditions and an approximant of degree $d = 24$.

5.2. *High friction limit of the Langevin equation*

Our second example is more fundamental in terms of physical relevance. As we have seen balanced averaging preserves the structure of the Langevin equation, including its statistical equilibrium properties. However there may be situations in which the structure-preservation turns out be very subtle. An interesting scenario in this respect is the overdamped limit of the Langevin equation that is also known by the name of inertia-less Langevin or Smoluchowski equation. The following result is due to Nelson [38].

Theorem 5.1 Let $(Q_t^\mu, V_t^\mu) \subset \mathbf{R}^n \times \mathbf{R}^n$ be the solution of

$$\begin{aligned} dQ_t^\mu &= V_t^\mu dt \\ \mu dV_t^\mu &= -\left(\nabla\varphi(Q_t^\mu) + \gamma V_t^\mu\right) dt + \sigma dW_t \end{aligned}$$

where $\varphi\colon \mathbf{R}^n \to \mathbf{R}$ is a smooth potential and friction and noise coefficients satisfy the fluctuation-dissipation relation $2\gamma = \sigma\sigma^T$. Then, as $\mu \to 0$, the process Q_t^μ converges with probability one to a diffusion process Q_t that is the solution of

$$\gamma dQ_t = -\nabla\varphi(Q_t)dt + \sigma dW_t\,.$$

The above limit is an example of a model reduction procedure, in which the reduced equations have a genuinely different structure (second-order vs. first-order). Nonetheless we can interpret the above result nicely in terms of the balancing method as we shall illustrate with a simple example. For $(Q_t^\mu, P_t^\mu) \in \mathbf{R} \times \mathbf{R}$, consider the equation

$$\begin{aligned} dQ_t^\mu &= \frac{1}{\mu} P_t^\mu dt \\ dP_t^\mu &= -\left(Q_t^\mu + \frac{1}{\mu} P_t^\mu\right) dt + \sqrt{2} dW_t \\ Y_t^\mu &= Q_t^\mu \end{aligned} \tag{5.2}$$

that describes damped oscillations of a stochastic particle of mass μ. Up to rescaling of time, sending $\mu \to 0$ is equivalent to letting γ diverge while keeping the temperature constant, i.e., $\gamma, \sigma \to \infty$ with $\gamma/\sigma^2 = \text{const}$; for this reason the zero-inertia limit $\mu \to 0$ is often referred to as the *high-friction limit*. It is easy to see

that (5.2) is a Langevin system of the form (3.4) with the Hamiltonian

$$H(q,p) = \frac{1}{2\mu}p^2 + \frac{1}{2}q^2$$

The two Gramians are

$$K = \begin{pmatrix} 1 & 0 \\ 0 & \mu \end{pmatrix}, \quad W_o = \frac{1}{2}\begin{pmatrix} 1+\mu & 1 \\ 1 & 1 \end{pmatrix}$$

with corresponding Hankel singular values $\sigma_1 \sim 1$ and $\sigma_2 \sim \mu$ for $\mu \to 0$. After balancing and averaging over the low energy mode and sending $\mu \to 0$ we obtain

$$\begin{aligned} dZ_t &= -Z_t dt + \sqrt{2} dW_t \\ Y_t &= Z_t \,. \end{aligned}$$

It is interesting to note that the dominant balanced variable (i.e., the one corresponding to σ_1) is of the form $z^\mu = q + p + \mathcal{O}(\mu)$, that is, the dominant balanced variable z is not just q as one might naively expect. Nonetheless the observed process converges in expectation, $\mathbf{E}[|Y_t^\mu - Y_t|] \to 0$ uniformly on $[0,T]$, as Theorem 4.1 asserts.

Appendix A. Derivation of the linear Langevin equation

We shall briefly sketch the derivation of a one-dimensional linear Langevin equation from deterministic Newtonian dynamics. The derivation goes as follows: consider a chain of $n+1$ identical particles of unit mass $m = 1$ that are connected by identical springs with spring constant $k = 1$. Letting q_i denote the elongation of the i-th particle from its rest position, Newton's equations for the chain read

$$\ddot{q}_i = q_{i+1} - 2q_i + q_{i-1} \,, \quad i = 0, \dots n$$

where we impose the Dirichlet boundary conditions, $q_{-1} = 0$, at the left end of the chain and Neumann conditions, $q_{n+1} = q_n$, at the other end.

Now let h be the equilibrium distance between neighbouring particles so that, setting $h = 1/n$, the total length of the chain becomes $l = 1$. Introducing the scaled time $\tau = t/h$ we recover the discrete wave equation

$$q_i'' = \frac{q_{i+1} - 2q_i + q_{i-1}}{h^2} \,, \quad i = 0, \dots n$$

with the notation $q' = dq/d\tau$. It is known that, as $n \to \infty$ and $h = 1/n$, the last equation converges to the continuous wave equation on an interval of length one with its left end fixed (Dirichlet b.c.) and the right end loose (Neumann b.c.).

Exploiting the linearity of the system, the equation for $q = (q_1, \dots q_n)$ can be solved analytically (e.g., by using Laplace transform). Plugging the result into the solution for the distinguished particle and integrating by parts once, we obtain a closed Volterra integro-differential equation for $Q = q_0$, namely,

$$Q''(\tau) + \int_0^\tau \gamma_n(\tau - \sigma) Q'(\sigma)\, d\sigma + k_n Q(\tau) = f_n(\tau) \tag{A1}$$

where $k_n \equiv 2/3$ and

$$\gamma_n(\tau) = \sum_{i=1}^{n} \alpha_i \cos(\omega_i \tau)\,, \quad f_n(\tau) = \sum_{i=1}^{n} \left(\eta_i \sin(\omega_i \tau) + \zeta_i \cos(\omega_i \tau)\right) \tag{A2}$$

with coefficients α_i and ω_i that can be explicitly computed, given the eigenvalues of the discrete wave equation, and $\eta_i = \eta_i(q(0), \dot{q}(0))$ and $\zeta_i = \zeta_i(q(0), \dot{q}(0))$ depending linearly on the initial conditions of the unresolved particles. Provided that the initial values for $q(0)$ and $q'(0)$ are known, equation (A1) is still deterministic—no randomness whatsoever.

The randomness comes into play, when we assume that the $q(0), q'(0)$ are independent and identically distributed Gaussian random variables with mean zero and unit variance in which case also f_n becomes a stationary Gaussian random process with mean zero and autocorrelation function

$$\mathbf{E}[f_n(\tau) f_n(\sigma)] = \gamma_n(\tau - \sigma)\,.$$

Note that the last equation is nothing but the fluctuation-dissipation (FD) relation for our problem. Finally, we may take the limit $n \to \infty$ in case of which γ_n becomes the Fourier cosine transform of Dirac's delta function which implies that f_n converges to a white noise process in the distributional sense, i.e., f_n becomes delta-correlated. This last statement can be rephrased by saying that f_n becomes the Karhunen-Loéve expansion of the white noise process—therefore, by the FD relation, the memory kernel converges to a Dirac delta at $\tau = 0$.

Hence in the thermodynamic limit $n \to \infty$, $h = 1/n$, the motion of the distinguished particle will be governed by a Langevin equation of the form

$$Q''(\tau) + \gamma Q'(\tau) + kQ(\tau) = \sqrt{\gamma}\xi(\tau) \tag{A3}$$

with ξ being a stationary white noise process.

The model of a particle coupled to a wave is known as Lamb's problem [23]. We are aware of the "hand-waviness" of our derivation, but it may serve the reader to understand that the Langevin equation is a plausible model for, e.g., linear materials that are surrounded by a heat bath (cf. [24]). For related studies of the Kac-Zwanzig heat bath model that gives rise to an intrinsically non-Markovian Langevin equation, the reader may consult the work [25].

Appendix B. The Support Theorem

Many properties of stochastic differential equations can be studied in terms of the corresponding control system. Consider the stochastic differential equation

$$dX_t = AX_t dt + BdW_t\,, \quad X_0 = x \tag{B1}$$

on $\mathbf{X} \subseteq \mathbf{R}^d$ with the associated control system

$$\dot{x}(t) = Ax(t) + Bu(t)\,, \quad x(0) = x \tag{B2}$$

Now let $V = C([0,T])$ be the space of continuous functions φ assuming values on $\mathbf{X} \subseteq \mathbf{R}^d$ and define $V_x = \{\varphi \in V : \varphi(0) = x\}$ to be the set of smooth curves starting in x. Let further X_t, $t \in [0,T]$ be the unique solution of the stochastic differential

equation (B1); obviously $X_t \in V_x$, and we may define the support of the diffusion process X_t as the smallest closed subset $U_x \subset V_x$ for which

$$\mathbf{P}\left[X_t \in U_x\right] = 1$$

with $\mathbf{P}[\cdot]$ being the probability measure on the path space V_x that is induced by the Brownian motion W_t. A typical question regarding the control system (B2) would be the following: given a control $u : [0, T] \to \mathbf{R}^m$ that is piecewise constant, which states can be reached in finite time T. In particular the space of integrable step functions is dense in L^2, so we can approximate any L^2 control by a series of step functions. We call a state $x' \in \mathbf{X}$ *reachable* if we can find a control u such that $x(0) = x$ and $x(T) = x'$ and we denote by $A_T(x)$ the set of states that are reachable within time T. The famous *Support Theorem* by Stroock and Varadhan [26] bridges the gap between the stochastic differential equation (B1) and its deterministic counterpart (B2).

Theorem B.1 Let $\phi_x(t; u)$ denote the solution of the controlled differential equation (B2). Then the support of the diffusion process X_t is given by

$$U_x = \overline{\{\phi_x(t; u) \colon u \in L(\mathbf{R}, \mathbf{R}^m)\}}$$

where $L(\mathbf{R}, \mathbf{R}^m)$ is the space of piecewise constant functions with values in $\mathbf{R}^m$.

It is a straight consequence of the last theorem that the reachable subspace of the stochastic differential equation is given by the set of states that can be reached using piecewise constant control input.

Corollary B.1 Let $\operatorname{supp}\mu \subseteq \mathbf{X}$ denote the support of a measure μ on $\mathbf{X}$. Then

$$\operatorname{supp} P_T(x, \cdot) = \overline{A_T(x)}\,, \quad T > 0\,.$$

where $P_t(x, C) = \mathbf{P}[X_t \in C \,|\, X_0 = x]$ with $C \subset \mathbf{X}$ being any open set denotes the transition probability of the Markov process X_t.

For linear systems such as (B2) complete controllability is guaranteed by Kalman's rank condition, i.e., $\operatorname{rank}(B|\, AB|\, A^2B|\ldots) = d$. By the support theorem, this implies that for all $t > 0$ the transition probability $P_t(\cdot, dy)$ of X_t has a smooth density $\rho_t(\cdot, y)$ with full topological support $\overline{A_t(\cdot)} = \mathbf{X}$. As a consequence, X_t is ergodic with respect to its invariant distribution $\mathcal{N}(0, W_c)$ with W_c being the infinite-time controllability Gramian (rate Gramian, respectively) of (B2).

Appendix C. The rate function

Lemma C.1 The rate function of the Langevin equation (3.4) is given by

$$L_r(x) = x \cdot K_T^{-1} x\,,$$

with $K_T = \operatorname{cov}(X_T)$ being the covariance matrix of the process X_t at time $t = T$.

Proof: We start by revisiting the well-known property of linear control systems to have a quadratic controllability function, and then show that it can be expressed in terms of the covariance matrix. Regarding the first, let $u \in L^2[0, T]$ be such that

$\phi_0(T;u) = x$ and consider the linear mapping $f\colon L^2[0,T] \to \mathbf{X}$ defined by

$$fu = \sqrt{\epsilon} \int_0^T \exp(A(T-s))Bu(s)\,ds\,.$$

By construction, we have $x = fu$. The adjoint map $f^* : \mathbf{X} \to L^2[0,T]$ is defined by means of the inner products

$$(f^*x, u)_{L^2[0,T]} = x \cdot fu$$

with the obvious generalization of the L^2 scalar product $(\cdot,\cdot)$ to vector-valued functions. Hence

$$(f^*x)\,(t) = \sqrt{\epsilon} B^T \exp(A^T(T-t))x$$

is an admissible control, i.e., the process with control $u = f^*x$ reaches x at time T. By ergodicity (i.e., complete controllability) the map f is onto which implies that $ff^* : \mathbf{X} \to \mathbf{X}$ is invertible. Now consider *any* admissible u with $x = fu$. The optimal such u is obtained by minimizing the L^2 norm $\|u\|^2_{[0,T]}$ subject to the constraint $x = fu$. The solution to this problem is provided by Hilbert's projection theorem [39], viz.,

$$u_* = f^*\,(ff^*)^{-1}\,x\,.$$

Obviously $u_* \in L^2[0,T]$ and, by the definition of the rate function, we obtain

$$L_r(x) = x \cdot (ff^*)^{-1}\,x$$

which completes the first part of the proof. Finally the assertion follows upon comparing ff^* to the expression (2.7) for the covariance matrix of X_T, namely,

$$\mathrm{cov}(X_T) = \epsilon \int_0^T \exp(As)BB^T \exp(A^T s)\,ds\,.$$

This concludes the proof. □

Appendix D. Numerical issues: proper orthogonal decomposition

The argument from Section C establishes a relation between controllability and the covariance matrix of a stable linear stochastic differential equation, and we may further exploit this correspondence so as to compute controllability and observability Gramians without solving Lyapunov equations.

Given any discrete realization $\{X_0, X_1, \ldots\}$ of (3.4) with arbitrary initial value $X_0 = x$, we define the empirical covariance matrix by

$$K_N = \frac{1}{N} \sum_{i=0}^{N-1} \left(X_i - \bar{X}_N\right)\left(X_i - \bar{X}_N\right)^T\,,$$

where

$$\bar{X}_N = \frac{1}{N}\sum_{i=0}^{N-1} X_i\,.$$

By stability of the Langevin process and the law of large numbers (i.e., ergodicity) we have $K_N \to K$ as $N \to \infty$ with probability one for almost all initial conditions X_0. But as the covariance matrix for $N \to \infty$ equals the controllability Gramian W_c—the reader should compare the respective Lyapunov equations—, we have just computed the controllability Gramian for (3.4). Conversely, we may compute the observability Gramian W_o from a sufficiently long realization of the adjoint stochastic system (complete observability assumed), which is numerically feasible, even if the system's dimension is too high to solve the corresponding Lyapunov equations.

The situation is even easier if the fluctuation-dissipation relation $2D = BB^T$ is in force. In this case any (sufficiently long) discrete trajectory $\{X_0, X_1, \ldots\}$ is distributed according to the equilibrium distribution $\rho_\infty \propto \exp(-H/\epsilon)$. In other words, the distribution of sample points $X_0, X_1, \ldots$ is Gaussian with mean zero and covariance given by $K = \epsilon E^{-1}$ where $E = \nabla^2 H$ is the constant Hessian of H. Consequently the rate or controllability Gramian for an equilibrium system, i.e., with coefficients satisfying $2D = BB^T$, is simply given by the inverse Hessian of the Hamiltonian.

Remark The empirical covariance matrix is the chief ingredient for computing low rank approximants of a given data set. For $\{X_0, X_1, \ldots, X_{N-1}\}$, the optimal rank-$d$ approximation

$$\min_{\Pi} \sum_{i=0}^{N-1} |X_i - \Pi X_i|^2 \quad \text{s.t.} \quad \Pi^2 = \Pi\,,\ \operatorname{rank}\Pi = d$$

is obtained by choosing Π to be the orthogonal projection onto the first d eigenvectors of K_N. Upon replacing the Euclidean inner product in the last equation by the Gramian-weighted one $|x|_o = \sqrt{\langle W_o x, x\rangle}$ and letting $N \to \infty$, the projection method recovers balanced truncation as has been pointed out in [40].

References

[1] G.W. Ford, M. Kac, and P. Mazur. *Statistical Mechanics of Assemblies of Coupled Oscillators.* J. Math. Phys. 6 (1965), pp. 504–516.
[2] R. Kubo. *The fluctuation-dissipation theorem.* Rep. Prog. Phys. 29 (1966), pp. 255–284.
[3] M. Allen and D. Tildesley. *Computer Simulations of Liquids.* Clarendon Press, 2005.
[4] E.G. Karpov, H.S. Park, and Wing Kam Liu. *A phonon heat bath approach for the atomistic and multiscale simulation of solids.* Int. J. Numer. Meth. Engng. 70 (2007) pp. 351–378.
[5] X. Li. *A coarse-grained molecular dynamics model for crystalline solids.* Int. J. Numer. Meth. Engng. 83 (2010), pp. 986–997.
[6] C.T. Mullis and R.A. Roberts. *Synthesis of minimum roundoff noise fixed point digital filters.* IEEE Trans. Circuits Systems 23 (1976), pp. 551–562.
[7] B.C. Moore. *Principal component analysis in linear systems: controllability, observability, and model reduction.* IEEE Trans. Auto. Contr. AC-26 (1981), pp. 17–32.
[8] K. Zhou, J.C. Doyle and K. Glover. *Robust and Optimal Control.* Prentice Hall, 1998.
[9] K. Glover. *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds.* Int. J. Control 39 (1984), pp. 1115–1193.
[10] A. Einstein. *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen.* Ann. Phys. (Berlin) 17 (1905), pp. 549–560.
[11] M. Smoluchowski. *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen.* Ann. Phys. (Berlin) 21 (1906), pp. 756–780.

[12] A.J. Chorin, O.H. Hald, and R. Kupferman. *Optimal prediction with memory.* Physica D 166 (2002), pp. 239–257.
[13] C. Hartmann. *Balancing of dissipative Hamiltonian systems.* Proc. MATHMOD 09, ARGESIM Report 35 (2009), pp. 1244–1255.
[14] B. Øksendal. *Stochastic differential equations: an introduction with applications.* Springer, 2003.
[15] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, 1993.
[16] J.M.A. Scherpen. *Balancing for nonlinear systems.* Syst. Control Lett. 21 (1993), pp. 143–153.
[17] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes.* North-Holland, 1981.
[18] J. Zabczyk. *Controllability of stochastic linear systems.* Syst. Control Lett. 1 (1981), pp. 25–31.
[19] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer, 1998.
[20] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems.* Springer, 1984.
[21] C. Hartmann, V.-M. Vulcanov, and Ch. Schütte. *Balanced truncation of linear second-order systems: a Hamiltonian approach.* Multiscale Model. Simul. 8 (2010), pp. 1348–1367.
[22] C. Hartmann and Ch. Schütte. *Balancing of partially-observed stochastic differential equations.* 47th IEEE Conference on Decisions and Control (2008), pp. 4867–4872.
[23] H. Lamb. *On a peculiarity of the wave-system due to the free vibrations of a nucleus in an extended medium.* Proc. London Math. Soc. s1-32 (1900), pp. 208–213.
[24] A.I. Komech and A.E. Merzon. *Scattering in the nonlinear Lamb system.* Phys. Lett. A 373 (2009), pp. 1005-1010.
[25] R. Kupferman. *Fractional kinetics in Kac-Zwanzig heat bath models.* J. Stat. Phys. 114 (2004), pp. 291–326.
[26] D.W. Stroock and S.R.S. Varadhan. *On the support of diffusion processes with applications to the strong maximum principle.* Proc. Sixth Berkeley Symp. Math. Statist. Prob. 3 (1972), pp. 333–359.
[27] T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide.* Springer, 2002.
[28] K. Ichihara and H. Kunita. *A classification of second order degenerate elliptic operators and its probabilistic characterization.* Z. Wahrscheinlichkeitstheorie verw. Gebiete 30 (1974), pp. 235–254.
[29] Y. Liu and B.D.O. Anderson. *Singular perturbation approximation of balanced systems.* Int. J. Control 50 (1989), pp. 1379–1405.
[30] A. Yu Veretennikov. *On rate of mixing and the averaging principle for hypoelliptic stochastic differential equations.* Math. USSR Izv. 33 (1989), pp. 221–231.
[31] E. Pardoux and A. Yu Veretennikov. *On the Poisson equation and diffusion approximation 3.* Ann. Prob. 33 (2005), pp. 1111–1133.
[32] E. Zeidler. *Applied Functional Analysis.* Springer, 1995.
[33] G.A. Pavliotis and A.M. Stuart. *Multiscale Methods: Averaging and Homogenization.* Springer, 2008.
[34] S.N. Ethier and T.G. Kurtz. *Markov Processes, Characterization and Convergence.* Wiley, 1986.
[35] A.J. Chorin, O.H. Hald, and R. Kupferman. *Optimal prediction and the Mori-Zwanzig representation of irreversible processes.* Proc. Natl. Acad. Sci. USA 97 (2000), pp. 2969–2973.
[36] Y. Chahlaoui and P. Van Dooren. *Benchmark examples for model reduction of linear time invariant dynamical systems.* In: P. Benner, D.C. Sorensen and V. Mehrmann, (eds). *Dimension Reduction of Large-Scale Systems. LNCSE 45*, pp. 379–392, Springer, 2005.
[37] X. Li and W. E. *Variational boundary conditions for molecular dynamics simulations of crystalline solids at finite temperature: Treatment of the thermal bath.* Phys. Rev. B 76 (2007), 104107.
[38] E. Nelson. *Dynamical Theories of Brownian Motion.* Princeton University Press, 1967.
[39] D.G. Luenberger. *Optimization by Vector Space Methods.* Wiley, 1997.
[40] C.W. Rowley. *Model Reduction for Fluids, Using Balanced Proper Orthogonal Decomposition.* Int. J. Bifurcat. Chaos 15 (2005), pp. 997–1013.