Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Balancing of partially-observed stochastic differential equations

Carsten Hartmann[1,2] and Christof Schütte[1]

Abstract— **We study Balanced Truncation for stochastic differential equations. In doing so, we adopt ideas from large deviations theory and discuss notions of controllability and obervability for dissipative Hamiltonian systems with degenerate noise term, also known as Langevin equations. For partially-observed Langevin equations, we illustrate model reduction by balanced truncation with an example from molecular dynamics and discuss aspects of structure-preservation.**

I. Introduction

Balanced Truncation is a rational and well established tool to reduce the dimension of a controlled linear differential equations [1]. Other, data-based approaches involve best-approximations by means of Proper Orthogonal Decomposition or Principal Component Analysis. Latter, in particular, is very popular in molecular dynamics as it is believed that the modes of highest variance carry important information, and indicate, e.g., conformational changes in a molecule [2]. The governing equations of molecular dynamics often take the form of stochastic differential equations that have the form of controlled systems, in which the smooth control variables have been replaced by suitable noise processes.

A class of stochastic differential equations that is of specific interest, e.g., in molecular dynamics, is given by *hypoelliptic* diffusion processes. Hypoelliptic diffusions entail certain dissipative Hamiltonian systems in which the white noise acts only on the momenta and which models dissipative dynamics in a heat bath. A representative of this class is the stochastic Langevin equation for which we study Balanced Truncation. For this purpose we employ a large deviations principle that allows for relating the sample paths of the white noise process to a smooth control variable under certain circumstances. This connection is not new indeed, and a variety of large deviations problems boil down to control arguments. A prominent example is, e.g., the Support Theorem of Stroock and Varadhan [3]; see also [4], [5].

We do not claim originality for the use of a large deviations principle in Section III for studying controllability of stochastic differential equations, nor do we claim complete mathematical rigour. However, our analysis reveals interesting relations between Balanced Truncation and empirical state-space decomposition methods such as the Principal Component Analysis which may have useful algorithmic implications; see Section IV. Furthermore, applying Balanced Truncation to stochastic differential equations brings up interesting mathematical questions, for example, if the negligible components are noisy and ought to be considered as a probability distribution rather than a point-wise process.

[1]Institut für Mathematik, Freie Universität Berlin, 14195 Berlin, Germany
[2] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

Another intriguing aspect that is related to structure-preservation is taken up in Section IV: On the one hand it has been shown recently [6] that Balanced Truncation applied to linear port-controlled dissipative Hamiltonian system preserves stability plus the Hamiltonian structure. On the other hand it is well known that, in the limit of small masses or high friction (dissipation), the Langevin process converges to a purely diffusive, inertia-less motion ("diffusive limit") that is described by a genuine first-order equation for the configurations. We discuss, when it may be advisable to relax structure-preservation in favour of physical considerations.

The article concludes in Section V with an example from molecular dynamics.

II. Stochastic Langevin equation

A frequently used model for dissipative Newtonian dynamics in a heat bath is the stochastic Langevin equation

$$M\ddot{q}(t) + \gamma\dot{q}(t) + \nabla V(q(t)) = \xi(t)\,, \tag{1}$$

where $q \in Q \subseteq \mathbf{R}^n$, and $\xi(t) \in \mathbf{R}^n$ denotes a Gaussian white noise process with covariance matrix

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

that is chosen so as to balance the energy dissipation due to the viscous friction. Here and in the following we shall use the symbol $\mathbf{E}(\cdot)$ to denote the expectation of a (measurable) stochastic process over all its possible realizations. Both mass and friction matrices $M, \gamma \in \mathbf{R}^{n\times n}$ are symmetric and positive definite.

Equation (1) is an instance of a dissipative Hamiltonian system that is driven by noise. Omitting the free variable t in what follows, we shall consider systems of the form[1]

$$\dot{x} \quad = (J - D(x))\,\nabla H(x) + S(x)\dot{W}\,, \tag{2}$$

where $x \in P$, $P = Q \times \mathbf{R}^n$, H is the system's Hamiltonian, and W denotes standard Brownian motion in $\mathbf{R}^n$. The matrix $J = -J^T$ is the canonical skew-symmetric matrix

$$J = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix},$$

and

$$D = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma \end{pmatrix}, \quad S = \begin{pmatrix} \mathbf{0} \\ \sigma \end{pmatrix}.$$

with σ, γ being invertible $n \times n$ matrices. Note that the noise acts only on the generalized momentum variables

[1]Strictly speaking, the notation $\dot{W}(t)$ does not make sense as the paths of the Brownian motion $W(t)$ are nowhere differentiable. However we use this notation for the sake of convenience and point out that the second equation has to be understood in the integral sense; see, e.g., [7].

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

$(x_{n+1}, \ldots, x_{2n})$. This entails that the usual Itô-Stratonovich dichotomy for stochastic differential equations vanishes, such that (2) behaves as an ordinary differential equation under point transformations [8]. From now on we call (2) a Langevin equation with (1) appearing as a special case.

If $\gamma = \gamma^T > 0$ the system is stable, i.e., for all stable critical points $e \in P$ of the energy function H, the eigenvalues of $(J - D(e))\nabla^2 H(e)$ lie in the open left half complex plane (cf. [9]). An important entity associated with (2) is its infinitesimal generator

$$L = \frac{1}{2} SS^T : \nabla^2 + (J - D)\, \nabla H \cdot \nabla \tag{3}$$

that generates the semigroup of solutions, where $A : B = \mathrm{tr}(A^T B)$ denotes the matrix inner product. If all coefficients in (3) are sufficiently smooth, then the operator L satisfies Hörmander's condition, i.e., L is hypoelliptic [10]. As a consequence, (2) has an ergodic invariant measure $d\rho$ with full topological support that is obtained as the solution of the stationary adjoint equation $L^*\rho = 0$. If moreover friction and noise coefficients satisfy the relation

$$2\gamma = \sigma\sigma^T \,,$$

then the solution of (2) is ergodic with respect to the Boltzmann measure $d\rho \propto \exp(-H)$, i.e., any infinitely long realization of the Langevin process samples ρ.

A. Linear Langevin equation

In this article we will consider only linear Langevin equations. That is, we study the Langevin equation (2) with

$$H(x) = \frac{1}{2} x^T E x \,, \quad E = E^T > 0$$

and the matrices D, S being constant. The reader may think of this linear Langevin system as, e.g., the Markovian limit of an optimal prediction strategy of a linearized Hamiltonian system or simply as a linearization of a nonlinear Langevin equation around a stable equilibrium (see the example below). With $D \in \mathbf{R}^{2n\times 2n}$ as given above, the drift matrix $A = (J - D)E$ is Hurwitz, i.e., all eigenvalues have strictly negative real part. Hence $x = 0$ is the unique stable equilibrium of the deterministic dynamics.

Given $X_0 = x$, the solution X_t, $t > 0$ of the linear Langevin is given by the stochastic integral [7]

$$X_t = x + \int_0^t AX_t\, dt + \int_0^t S\, dW_t \,.$$

As is straightforward to verify using Itô's formula, the last equation is equivalently expressed as

$$X_t = e^{At}x + \int_0^t e^{A(t-s)} S\, dW_s \,. \tag{4}$$

We stress that some of the considerations in this article carry over to the nonlinear case as well, such as the large deviations principle — at least on a formal level. However the analysis may differ in the details and we feel that most of the ideas can be stated in a more transparent way for the linear problem; so we shall stick to this case here. For balancing of nonlinear control systems we refer to, e.g., [11].

III. A LARGE DEVIATIONS PRINCIPLE

The Hörmander property of the Langevin equation essentially guarantees that the noise in the system spreads over the full phase space as the system evolves in time which implies that the system (2) is completely controllable in the sense of control theory [12]. A notion of stochastic controllability of is provided by large deviations theory; see, e.g., [5], [13]. Let us briefly recall some basic concepts: For $\epsilon > 0$, we consider the family of stochastic Langevin equations

$$\dot{x} \quad = (J - D)\, \nabla H(x) + \sqrt{\epsilon} S\dot{W} \tag{5}$$

with $x(0) = x$ that have the solutions

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

again with the abbreviation $A = (J-D)\nabla^2 H(x)$. Let further $C([0,T])$ be the space of continuous functions taking values in $P = Q \times \mathbf{R}^n$, and $H^1([0,T])$ the space of absolutely continuous functions with values in P and square-integrable derivatives. We assume that for all $t < \infty$ the solutions X_t^ϵ of (5) are continuous, and we denote by $\mathbf{P}_x^\epsilon$ the probability measure induced on $C([0,T])$ by the processes X_t^ϵ starting at $X_0^\epsilon = x$. We introduce the rate function

$$I_x(f) = \inf_{u\in H^1,\, f = F(u)} \frac{1}{2} \int_0^T |\dot{u}(t)|^2\, dt \tag{6}$$

with $u(0) = 0$ and $f = F(u)$, where $F : H^1([0,T]) \to C([0,T])$ is given by

$$F(u)(t) = e^{At}x + \int_0^t e^{A(t-s)} S\dot{u}(s)\, ds \,. \tag{7}$$

We declare that $I_x(f) = \infty$ if there is no $u \in H^1$ such that $f = F(u)$. The idea of the thus defined rate function is to replace W_t by its polygonal approximation. Due to the scaling of the noise term in (5) the approximation error $\|f(t) - X_t^\epsilon\|$ vanishes as ϵ goes to zero, so we may expect that the smooth approximation does not influence the final result. We state the large deviations principle (LDP) that can be found in [13]:

Proposition 3.1: Let $X_t^\epsilon : [0,T] \to P$ with $T < \infty$ be the random solution of (5) with $X_0^\epsilon = x$ fixed. Then the rate function $I_x(\cdot)$ is lower semi-continuous, and X_t^ϵ satisfies a LDP. That is, for each open set $\mathcal{A} \subset C([0,\infty])$

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{P}_x^\epsilon \left(X_{(0,T]}^\epsilon \in \mathcal{A} \right) \geq - \inf_{f\in\mathcal{A}} I_x(f) \,,$$

whereas for each closed set $\mathcal{B} \subset C([0,\infty])$

$$\limsup_{\epsilon\to 0} \epsilon \log \mathbf{P}_x^\epsilon \left(X_{(0,T]}^\epsilon \in \mathcal{B} \right) \leq - \inf_{f\in\mathcal{B}} I_x(f) \,.$$

Notice that the infinum in the rate function $I_x(\cdot)$ is taken over functions $u \in H^1$ that are smoother than the ordinary realizations of the Brownian motion. Roughly speaking, the LDP makes an assertion concerning the probability of smooth solutions of the Langevin equation in the limit of weak noise. In other words, we may treat the white noise process $\dot{W}(t)$ in (2) like a smooth control variable, provided that the matrix S is sufficiently small in any appropriate matrix norm. For details we refer to [3].

A. Controllability and observability

The above considerations suggest that a controllability function for (2) can be defined in the following fashion:

$$L_c(z) = \inf_{f(T)=z} I_0(f)\,, \tag{8}$$

where the rate function $I_{x=0}(f)$ is given by (6) and $f = F(u)$ is defined according to (7) with $x = 0$. The function L_c describes the minimal energy that is needed for the process to reach $z \in P$ after time $t = T$, when it was started at $x = 0$ at time $t = 0$. Note that, by definition, $L_c = \infty$ when $u \notin H^1$. Now we can state:

Proposition 3.2: Consider the Langevin equation (2) with constant $D, S \in \mathbf{R}^{2n\times 2n}$ and quadratic Hamiltonian H, such that $A = (J - D)\nabla^2 H(x)$ is Hurwitz. Then, for all $z \neq 0$, the controllability function $L_c(z) > 0$ is given by

$$L_c(z) = \frac{1}{2} z^T K_T^{-1} z\,, \tag{9}$$

where $K_T = \mathrm{cov}(X_T)$ is the covariance matrix of the unscaled Langevin process (4) 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 $f = F(u)$ with $u \in H^1$ be such that $f(T) = z$ and consider the linear mapping $C\colon H^1([0,T]) \to P$ defined by

$$Cu = \int_0^T e^{A(T-s)} S\dot{u}(s)\, ds\,.$$

By construction, we have $z = Cu$. The adjoint map $C^* : P \to H^1([0,T])$ is defined by means of the inner products

$$\langle z, Cu\rangle_{\mathbf{R}^{2n}} = \langle C^* z, u\rangle_{H^1}\ .$$

Hence,

$$(C^* z)(t) = S^T e^{A^T(T-t)} z$$

is an admissible approximation of $\dot{W}$, such that the process reaches z at time T. Since A is Hurwirtz and (2) is completely controllable the map C is onto. Consequently, the map $CC^* : P \to P$ is invertible. Now consider any $u \in H^1$ with $z = Cu$. The optimal such u is given by minimizing $\|u\|^2_{H^1} = \|\dot{u}\|^2_{L^2}$ subject to the constraint $z = Cu$. The solution to this problem is given by the projection theorem:

$$\dot{u}_0 = C^* (CC^*)^{-1} z\,.$$

Obviously $u_0 \in H^1$ and we can use (8) together with the definition of the rate function to obtain

$$L_c(z) = \frac{1}{2} z^T (CC^*)^{-1} z\,.$$

This completes the first part of the proof. As for the identity $CC^* = \mathrm{cov}(X_T)$, consider the solution

$$X_t = \sqrt{\epsilon} \int_0^t e^{A(t-s)} S\, dW_s$$

of the Langevin equation (2) for the initial value $X_0 = 0$. Since $\mathbf{E}X_t = 0$, the covariance matrix turns out to be

$$\begin{aligned}\mathrm{cov}(X_t) &= \mathbf{E} X_t X_t^T \\ &= \mathbf{E}\left(\int_0^t e^{A(t-s)} S dW_s \int_0^t dW_s^T S^T e^{A^T(t-s)}\right)\end{aligned}$$

which, by using the Itô isometry, can be recast as

$$\begin{aligned}\mathrm{cov}(X_t) &= \int_0^t e^{A(t-s)} SS^T e^{A^T(t-s)}\, ds \\ &= \int_0^t e^{As} SS^T e^{A^T s}\, ds\,.\end{aligned}$$

The assertion follows upon noting that $CC^* = \mathrm{cov}(X_T)$. ■

As T goes to infinity, the Gramian CC^* converges to the equilibrium covariance matrix of the Langevin process that is well-defined for A being Hurwitz. Using integration by parts in the expression of $K_T = \mathrm{cov}(X_T)$ for $T \to \infty$, the equilibrium covariance K is the unique and symmetric positive-definite solution of the Lyapunov equation

$$AK + KA^T = -SS^T\,.$$

Observability of the stochastic Langevin equation can be established in the standard way as for any other controlled ordinary differential equation; see [1]. To this end, we write the observed Langevin equation in the common form of a port-controlled Hamiltonian system [14]

$$\begin{aligned}\dot{x} &= (J - D)\,\nabla H(x) + S\dot{W} \\ y &= R\nabla H(x)\,.\end{aligned} \tag{10}$$

We can now define the observability function by disregarding the noise contribution, viz.,

$$L_o(x) = \frac{1}{2}\int_0^T |y(t)|^2\, dt\,, \tag{11}$$

where $y(t) = R\nabla H(f(t))$ with $f = F(0)$ being the solution of the deterministic dynamics with initial condition $f(0) = x$. Since $f(t) = \exp(At)x$, the observability function reads

$$L_o(x) = \frac{1}{2} x^T O_T x\,,$$

where O_T denotes the finite-time observability Gramian

$$O_T = \int_0^T e^{A^T t} G^T G e^{At}\, dt\,, \quad G = R\nabla^2 H(x)\,.$$

Again, as we let T go to infinity, the observability Gramian can be represented as the unique positive-definite symmetric solution of the Lyapunov equation

$$A^T O + OA = -G^T G\,.$$

IV. Balanced Truncation

Controllability and observability Gramians both have suggestive physical interpretations: The controllability Gramian K measures to which extend states are excitable by the noise. Given two states x_1 and x_2 with $|x_1| = |x_2|$, x_1 is more sensitive than x_2, if $x_1^T K x_1 > x_2^T K x_2$. Conversely, disregarding the noise, x_1 is better to observe than x_2, if it produces a higher output energy, i.e., if $x_1^T O x_1 > x_2^T O x_2$.

Balancing consist in finding a coordinate transformation, such that states can be simultaneously ordered according to the excitability and their output energy. This is achieved by finding a transformation $x \mapsto Tx$ that makes both Gramians equal and diagonal, viz.,

$$T^{-1}KT^{-T} = T^T OT = \mathrm{diag}(\sigma_1, \ldots, \sigma_{2n})\,.$$

A basis result is that such a transformation exists, whenever the Gramians are positive definite. Moreover the Hankel singular values $\sigma_1, \ldots, \sigma_{2n}$ are invariant under coordinate changes. In the balanced coordinates those states that are least excitable also give the least output which legitimates to neglect them, provided the corresponding Hankel singular values are sufficiently small (truncation).

The linear Langevin equation (10) resembles a stable port-controlled Hamiltonian system. In [6] it has been demonstrated that constraining the system to the even-dimensional subspace $U \subset P$ of the most controllable and observable states results again in a stable port-Hamiltonian system. In point of fact, the problem reduces to imposing a holonomic constraint on a Hamiltonian system, either directly by restriction or by employing singular perturbation techniques. If the constraint $U \subset P$ is holonomic (i.e. integrable), imposing constraints for the stochastic Langevin equation works in the same manner as for any other Hamiltonian systems; we refer to [8] for the details regarding the constraint problem for stochastic differential equations.

The considerations in [6] carry over to the Langevin problem – at least partially. Given coordinates $z = (z_1, \ldots, z_{2k})$, $k < n$ on $U \subset P$, the truncated version of the Langevin equation (10) reads

$$\begin{aligned} \dot{z} &= (J_\mathrm{r} - D_\mathrm{r})\,\nabla H_\mathrm{r}(z) + S_\mathrm{r}\dot{W} \\ y &= R_\mathrm{r}\nabla H_\mathrm{r}(z)\,, \end{aligned} \tag{12}$$

with $J_\mathrm{r} - D_\mathrm{r} = V_1(J-D)V_1^T$, $S_\mathrm{r} = V_1 S$ and $R_\mathrm{r} = RV_1^T$, where V_1 contains the first $2k < 2n$ rows of $V = T^{-1}$. The reduced Hamiltonian is obtained simply by restriction, i.e.,

$$H_\mathrm{r} = \frac{1}{2} z^T E_\mathrm{r} z\,, \quad E_\mathrm{r} = T_1^T E T_1\,.$$

Here T_1 consists of the first $2k$ columns of T. The reduced system is stable, i.e., the matrix $A_\mathrm{r} = (J_\mathrm{r} - D_\mathrm{r})\nabla^2 H_\mathrm{r}(z)$ is Hurwitz. (Alternatively, one may replace E_r by the the Schur complement of the balanced matrix $\tilde{E} = T^T E T$ resulting also in a stable reduced system.) Moreover $J_\mathrm{r} = -J_\mathrm{r}^T$ is skew-symmetric as is easily seen. Note that, if that the original system satisfied the fluctuation-dissipation relation $2D = SS^T$, then the same is true for the truncated system.

Nonetheless some care is needed in the interpretation of the noise term, for the dimension of the Brownian motion $W(t) \in \mathbf{R}^n$ has not changed. Consequently it may happen that $2k < n$ which implies that the process is no longer Langevin, and it matters whether we regard the equation either as an Itô or an Stratonovich equation.

A final remark is in order. The usual singular perturbation argument of Balanced Truncation does not directly apply in the presence of unbounded white noise. If the noise is acting on the subsystem that corresponds to the smallest singular values contains noise, the dynamics does not contract to the controllable and observable subspace as the small Hankel singular values go to zero; cf. [6]. In fact, the variables become distribution-valued (in the sense of probability theory) which brings us into the realm of averaging techniques [15].

A. Empirical state-space decomposition

The argument from Section III establishes a relation between controllability of linear control systems and the covariance matrix of a stable linear stochastic differential equation. Given any (discrete) realization $\{X_0, X_1, \ldots\}$ of (2), 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 and ergodicity of the Langevin process we have $K_N \to K$ as $N \to \infty$ with probability one for almost all initial conditions $X_0 = x$. Likewise, we may compute the observability Gramian O from realizations of the adjoint system (complete observability assumed), which is numerically feasible, even if the system's dimension is too high so as to solve the corresponding Lyapunov equations.

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-$2k$ approximation

$$\min_{\Theta_k} \sum_{i=0}^{N-1} \|X_i - \Theta_k X_i\|^2 \quad \text{s.t.} \quad \Theta_k^2 = \Theta_k$$

is obtained by choosing Θ_k to be the orthogonal projection onto the first $2k$ eigenvectors of K_N. Upon replacing the Euclidean inner product in the last equation by the Gramian-weighted one, $\|x\|_O = \sqrt{\langle Ox, x\rangle}$ and letting $N \to \infty$, the projection method recovers the Balanced Truncation subspace as has been pointed out in [16]; cf. also [17].

B. Diffusive limit

Balanced Truncation as carried out in the just described way preserves the port-Hamiltonian structure of the Langevin equation, let alone the exact meaning of the noise term. However there may be situations in which structure-preservation may be relaxed in favour of a physical considerations.

An interesting object in this respect the diffusive limit of the Langevin equation that also known by the name of diffusive limit or Smoluchowski equation. The following remarkable result is due to Nelson [18]; we have adapted it so as to fit our framework.

Proposition 4.1: Let $(Q_t^\epsilon, P_t^\epsilon)$ denote the solutions of

$$\begin{aligned} \dot{q} &= \frac{\partial H^\epsilon}{\partial p} \\ \dot{p} &= -\frac{\partial H^\epsilon}{\partial q} - \gamma\frac{\partial H^\epsilon}{\partial p} + \sigma\dot{W}\,, \end{aligned}$$

where friction and noise coefficients satisfy $2\gamma = \sigma\sigma^T$, and H^ϵ is a family of Hamiltonians given by

$$H^\epsilon(q,p) = \frac{1}{2\epsilon} p^T M^{-1} p + \frac{1}{2} q^T L q \,.$$

Then, as $\epsilon \to 0$, the process Q_t^ϵ converges with probability one to a diffusion process Q_t^0 that is the solution of

$$\gamma \dot{q} = -Lq + \sigma \dot{W} \,. \tag{13}$$

The diffusive limit of the Langevin equation 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 Balanced Truncation as we shall illustrate with a simple example. For $x_1 \in \mathbf{R}$, consider the equation

$$\epsilon \ddot{x}_1 = -x_1 - \dot{x}_1 + \sqrt{2}\dot{W} \,, \tag{14}$$

that describes damped oscillations of a stochastic particle of mass ϵ. If we rescale the free variable according to $t \mapsto \epsilon t$ the last equation turns out to be equivalent to the system[2]

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\epsilon x_1 - x_2 + \sqrt{2\epsilon}\dot{W} \,. \end{aligned} \tag{15}$$

By standard perturbation arguments, we might guess that, as $\epsilon \to 0$, the dynamics degenerates to the system

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_2 \,, \end{aligned}$$

which implies $x_2 \approx \exp(-t)x_{2,0}$ and $x_1 \approx -x_2$ for $\epsilon \ll 1$. However we have to be careful in neglecting terms involving ϵ, for the white noise is unbounded and, hence, both $\epsilon\dot{x}_1$ and $\sqrt{2\epsilon}\dot{W}$ can be become arbitrarily large, no matter what ϵ is.

Now suppose that we observe only the position component $y = x_1$ of the system (15). As the noise amplitude is small for $\epsilon \ll 1$ the large deviations principle of Section III applies, thus we can study controllability and observability of the system. The two Gramians are easily found, viz.,

$$K = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} \,, \quad O = \frac{1}{2\epsilon} \begin{pmatrix} 1+\epsilon & 1 \\ 1 & 1 \end{pmatrix} \,.$$

The corresponding Hankel singular values are

$$\sigma_{1/2} = \frac{1}{2} \sqrt{\frac{1 + 2\epsilon \pm \sqrt{1+4\epsilon}}{\epsilon}}$$

yielding $\sigma_1 \sim 1/\sqrt{\epsilon}$ and $\sigma_2 \sim \sqrt{\epsilon}$ for $\epsilon \to 0$. Computing the balancing transformation and truncating the low energy modes for $\epsilon \ll 1$, we obtain the diffusion equation

$$\dot{\xi}_1 = -\epsilon \xi_1 + \sqrt{2\epsilon}\dot{W}$$

with the single balanced variable $\xi_1 = x_1$ as the reduced form of the scaled Langevin equation (15). If we scale back to the original time scale by $\epsilon t \mapsto t$, we find that x_1 in the original problem (14) is best described by the diffusion process X_t^0 that is the solution of

$$\dot{x}_1 = -x_1 + \sqrt{2}\dot{W} \,,$$

[2]Notice that the white noise scales according to $\dot{W}(t) \mapsto \sqrt{\epsilon}\dot{W}(t/\epsilon)$.

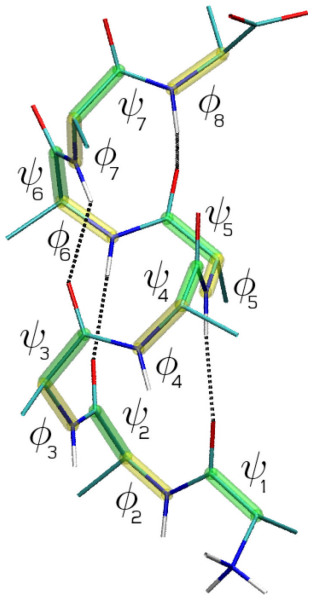

Fig. 1. Helical conformation of 8-alanine

which in fact is of the diffusive limit form (13). The last equation defines a Gaussian diffusion process X_t^0 with

$$\mathbf{E}X_t^0 = e^{-t}x_{1,0} \,, \quad \mathrm{cov}(X_t^0) = 1 - e^{-2t} \,.$$

V. Molecular dynamics

We want to study Balanced Truncation for a realistic molecular system: the backbone angle dynamics of 8-alanine in water at 300K. The corresponding Hamiltonian is clearly a nonlinear function in the dihedral angles and their conjugate momenta, so we linearize around a stable fixed point. Such a fixed point is, e.g., given by the stable helical conformation of the alanine molecule (see Fig. 1).

As the system's Hamiltonian is typically given in Cartesian coordinates rather than dihedral angles, we employ a data-based ansatz, whereupon the right-hand side of the Langevin equation is estimated from an observation time series consisting of 7 backbone angles pairs $q = (\phi_2, \psi_1, \ldots, \phi_8, \psi_7)$ and their angular velocities $\dot{q} = dq/dt$. The molecular dynamics simulation is performed with the GROMOS force field and implicit solvent with a 1fs time step and total length of 1ns. The angular velocities are obtained accurately via force field evaluations within the GROMOS package. The dynamics was restricted to the helical conformation such that the assumption of linear Langevin dynamics seems reasonable. The parameters of the respective linear Langevin equation

$$M\ddot{q} = -Lq - \gamma\dot{q} + \sigma\dot{W} \tag{16}$$

are estimated using the maximum-likelihood routine described in [19], assuming dissipation-fluctuation relation $2\gamma = \sigma\sigma^T$ to hold. From the optimal parameters (see Fig. 2) we then compute the Hankel singular values that are shown in Figure 3 where, as before, we have chosen the observable $y = q$. We see that the singular values are quickly decaying with three dominant ones.

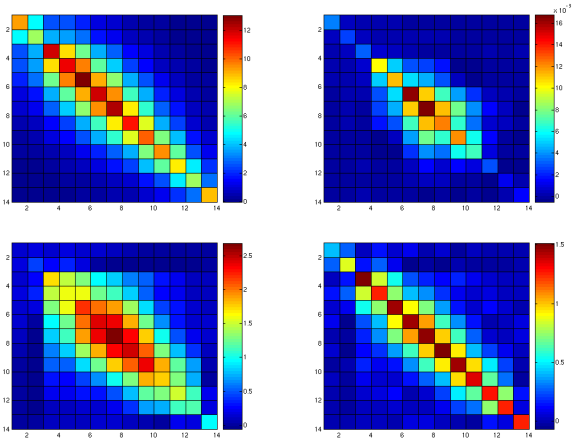

Fig. 2. Optimal model parameters of the linearized Langevin equation around the helical conformation. Upper panel: mass M and stiffness matrix L. Lower panel: friction and noise coefficients, γ and σ.

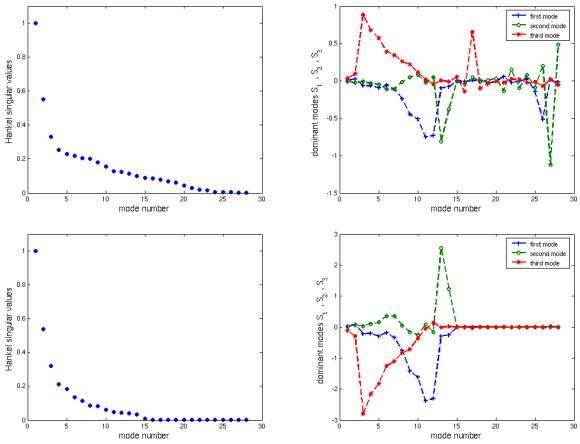

Fig. 3. Hankel singular values and the corresponding inverse balancing transformation. Upper panel: Hankel SV and the first three rows of T^{-1} for $\mu = 1$ (estimated parameters). Lower panel: Hankel SV and the first three rows of T^{-1} for $\mu = 100$ (the largest SV has been scaled to unity).

By looking at the corresponding transformation matrix (upper panel in Fig. 3) we see that the most pronounced modes are essentially configuration variables (i.e., angles), where the largest contributions to these modes come from the central dihedral angles of the alanine backbone. The last observation might probably be caused by lower mobility of the (implicit) solvent in the central regime of the peptide.

The finding that the most pronounced modes are essentially configuration variables is in agreement with our finding about the diffusive limit: as Figure 2 shows, mass and the friction matrix are about two orders of magnitude larger than the stiffnesses. In order to understand the effects of this difference in scaling let uns introduce

$$M = \frac{1}{\epsilon} M_0 \,, \quad \gamma = \frac{1}{\epsilon} \gamma_0 \,, \quad \sigma = \frac{1}{\sqrt{\epsilon}} \sigma_0 \,,$$

and define the "compressed" time $\tau = \epsilon t$, we see that the Langevin equation (16) is indeed of the form (14), namely,

$$\epsilon M_0 \frac{d^2 q}{d\tau^2} = -Lq - \gamma_0 \frac{dq}{d\tau} + \sigma_0 \dot{W} \,,$$

where M_0, L, γ, σ are all of order 1. To further emphasize the effect of scale separation we compare the computed Hankel singular values and the balancing transform for the estimated parameters M, L, γ, σ with the upscaled parameters $M \mapsto \mu M$ and $\gamma \mapsto \mu\gamma$ (accordingly: $\sigma \mapsto \sqrt{\mu}\sigma$), while keeping the stiffness matrix L fixed. The result for $\mu = 100$ is shown in the lower panel of Figure 3. It turns out that as μ increases, half of the Hankel singular values goes to zero (uniformly proportional to $1/\sqrt{\mu}$). In the same way the contribution of the momenta to the dominant modes disappears, and the dominant modes become purely configurational.

VI. Acknowledgements

Sincere thanks is given to Frank Noé for the simulation data and to Jochen Voß for mentioning large deviations. This work is supported by the DFG Research Center Matheon "Mathematics for Key Technologies" (FZT86) in Berlin.

References

[1] B.C. Moore. Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction. *IEEE Trans. Auto. Contr.* **AC-26**(1):17–32, 1981.

[2] T. Schlick. *Molecular Modeling and Simulation - An Interdisciplinary Guide.* Springer, 2002.

[3] 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**: 333–359, 1972.

[4] G.B. Arous and J.-D. Deuschel. The rate function of hypoelliptic diffusions. *Comm. Pure Appl. Math.* **47**:843–860, 1994.

[5] J. Feng. Martingale problems for large deviations of Markov processes. *Stochastic Process. Appl.* **81**:165–216. 1999.

[6] C. Hartmann, V.-M. Vulcanov, and Ch. Schütte. Balanced truncation of linear second-order systems: A Hamiltonian approach, *submitted to Mult. Mod. Sim.*, 2007.

[7] B. Øksendal. *Stochastic differential equations : an introduction with applications.* Springer, 2003.

[8] C. Hartmann. *Model Reduction in Classical Molecular Dynamics.* Dissertation thesis, Freie Universität Berlin, 2007, *available via http://www.diss.fu-berlin.de/2007/458.*

[9] J.H. Maddocks and M.L. Overton. Stability theory for dissipatively perturbed Hamiltonian systems. *Comm. Pure Appl. Math.* **48**:583–610, 1995.

[10] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[11] J.M.A. Scherpen. Balancing for nonlinear systems. *System &. Control Letters* **21**:143-153, 1993.

[12] A. Isidori. *Nonlinear Control Systems.* Springer, 1995.

[13] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer, 1998.

[14] A. van der Schaft. *L_2-Gain and Passivity Techniques in Nonlinear Control.* Springer, 1996.

[15] C. Hartmann. Structure-preserving model reduction of partially observed differential equations: molecular dynamics and beyond. *Oberwolfach Reports* **21**: 33–37, 2008.

[16] C.W. Rowley. Model Reduction for Fluids, Using Balanced Proper Orthogonal Decomposition. *Int. J. on Bifurcation and Chaos* **15**(3):997–1013, 2005.

[17] S. Lall, J.E. Marsden, and S. Glavaški. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. Robust Nonlin. Control* **12**: 519–535, 2002.

[18] E. Nelson. *Dynamical Theories of Brownian Motion.* Princeton University Press, 1967.

[19] I. Horenko and Ch. Schütte. Likelihood-based estimation of multidimensional Langevin models and its application to biomolecular dynamics. *To appear in Mult. Mod. Sim.*, 2008.