Adaptive Integration of Molecular Dynamics

ILLIA HORENKO,[1] MARTIN WEISER[2]
[1]*Freie Universität Berlin, Arnimallee 2, D-14195 Berlin, Germany*
[2]*Zuse Institute Berlin, Takustr. 7, D-14195 Berlin, Germany*

Received 2 August 2002; Accepted 5 June 2003

Abstract: This article presents a particle method framework for simulating molecular dynamics. For time integration, the implicit trapezoidal rule is employed, where an explicit predictor enables large time steps. Error estimators for both the temporal and spatial discretization are advocated, and facilitate a fully adaptive propagation. The framework is developed and exemplified in the context of the classical Liouville equation, where Gaussian phase-space packets are used as particles. Simplified variants are discussed briefly. The concept is illustrated by numerical examples for one-dimensional dynamics in double well potential.

 J Comput Chem 24: 1921–1929, 2003

Key words: Gaussian particle methods; adaptivity; error estimation; classical Liouville equation; molecular dynamics

Introduction

The foundation of mathematical descriptions of molecular dynamics is provided by quantum theory in the form of the time-dependent Schrödinger equation. This partial differential equation, however, is defined on function spaces with a dimension proportional to the number of atoms involved. Already for medium-size molecules, the curse of dimensionality leads to an exponentially growing computational cost of traditional grid discretization techniques based on finite differences, finite elements, or the Fourier transform.

As a remedy for the curse of dimensionality, two alternatives to traditional grid discretization techniques are available: *sparse grids* (cf. refs. 1 and 2) and *particle methods* (cf. refs. 3 and 4), which both scale reasonably well to medium dimensional problems. Sparse grids, however, are best suited for representing smooth densities with grid-aligned features.

Particle methods for simulating quantum mechanical systems tackle either the time-dependent Schrödinger equation directly,[5–8] or the Wigner-transformed Liouville–von Neumann equation, which casts the evolution into the phase space spanned by positions and momenta.[9–11] Approximations to the Liouville–von Neumann equation, for example, the classical Liouville equation, have also attracted interest.[12–15]

The Schrödinger equation and its reformulations and approximations can be discretized by particle methods with different particle shape functions. A common approach is to approximate the phase space distributions by collections of Dirac functional trajectories (cf. ref. 16). In this case, the dynamics is reduced to Newton's equations of motion, which are routinely solved in classical molecular dynamics simulations. The attractive simplicity of such a local particle base has, however, three major drawbacks. First, a Dirac function representation is hardly appropriate for problems where quantum effects, and hence, nonlocal effects in continuous distributions play an important role. In these situations, which include nonadiabatic population exchanges,[17–20] multidimensional potential energy surfaces with high barriers,[13] and nonclassical forces occurring in the "Bohmian" formulation of quantum mechanics,[21] approximating the continuous Wigner distributions by collections of smooth particle shape functions is far more appropriate. Second, the computation of correlations and overlap integrals is difficult if only point values are available. A continuous representation of the densities can be expected to simplify this task.[13] Third, a singular representation of continuous quantum-mechanical distribution functions makes error estimation difficult, and thus complicates the construction of spatially adaptive simulation algorithms.

Particle methods based on a superposition of Gaussian wave packets as introduced by Heller[6,22–25] have become popular, and inspired a number of related methods, for example, the multiple spawning method[17–19] and the multithreads method.[26–28] Quite often, the proposed algorithms rely on two simplifying assumptions, which have been discussed in detail by Sawada, Heather, Jackson, and Metiu:[7] (1) the independent particle approximation (IPA, also known as IGA—independent Gaussians approximation), which assumes that the particles can be propagated independently, and is thus closely connected to the flow-conservation property of the dynamics; and (2) the locally harmonic approxi-

Correspondence to: M. Weiser; e-mail: weiser@zib.de

Contract/grant sponsor: SFB 450: Ultraschnelle photoinduzierte Prozesse

mation (LHA), which assumes that the width of each Gaussian is smaller than the length over which the potential deviates significantly from a quadratic shape. Note that Dirac function representations as commonly realized in molecular dynamic codes rely on flow conservation, and hence, the IPA assumption. Both assumptions are sufficiently valid in a number of practically relevant situations for short simulation times, a fact that has been exploited by Heller to simulate such processes by a comparatively simple numerical scheme.

There are, however, several situations where neither IPA nor LHA are valid, for example, reduced models or nonconservative systems violate the IPA, whereas the LHA is in general violated for realistic potentials and propagation times. This motivated the development of algorithms that do not depend on these assumptions. The strategy proposed by Walkup et al.[29] and Prezhdo et al.[30] exploits the idea of Taylor expansions by using higher order derivatives of the potential for propagating the distribution function. However, in the case of real-life applications (assuming many important degrees of freedom for the underlying molecular system) the problem of calculation of this derivative becomes intractable. Sawada et al.[7] introduced the minimum error method (MEM), which is essentially the *method of lines,*[31] where the equations of motion are derived by a least squares approximation to the continuous evolution. The method is successfully applied to processes violating both LHA and IPA, but exhibits two weaknesses closely related to spatial adaptivity: (a) if the initially chosen number of Gaussians is too large, the equations of motion can become singular, and hence intractable in the course of the propagation; and (b) the number of Gaussians needed to represent the wave function later in time can substantially increase in time. A heuristic strategy for adapting the number of Gaussians based on monitoring eigenvalues of the overlap matrix has been sketched, but seems to be inefficient.

The present article addresses both weaknesses by advocating a Trapezoidal Rule Adaptive Integrator for Liouville type equations (TRAIL). In a theoretically backed way, the local approximation error is employed for creating new Gaussians where needed, and the subcondition number of the least-squares system matrix is exploited for removing the Gaussians that are no longer necessary for representing the continuous distribution. This *fully adaptive* scheme blends smoothly with the implicit trapezoidal rule used for propagation in time. In contrast to Wan and Schofield,[26–28] the selection of a suitable set of Gaussians aims at maintaining a user requested accuracy instead of a given computational effort per time step.

The method is exemplified and demonstrated in the simple context of the classical limit $\hbar \to 0$ of the Liouville–von Neumann equation, the *classical Liouville equation* (CLE), describing the dynamical behavior of a classical (quasi-)distribution function at constant energy in phase space. Note that the CLE simply transports density values along classical trajectories—a structural property that can be exploited for constructing efficient explicit methods. In contrast to the adaptive propagation scheme presented here, such methods cannot be extended to the quantum-classical Liouville equation or the Schrödinger equation, which do not exhibit a similar conservation property.

In the context of the CLE, we can utilize Monte Carlo approximation and explicit propagation methods for the CLE recently developed by Horenko, Schmidt, and Schütte[20] as building blocks in the proposed numerical scheme. Moreover, the presented method can easily be integrated into existing MD codes, thus providing the advantages of adaptive propagation and error-estimation for a wide spectrum of practical applications.

The remainder of the article is organized as follows. The next section is devoted to the time integration of the CLE dynamics by both implicit and explicit integrators, together with error control by adapting the time-step size. Then, the discretization of Wigner distributions by collections of Gaussians is described, and the adaptive refinement and coarsening of the approximation is presented. Finally, the Numerical Example section contains numerical examples.

Time Integration of the CLE Dynamics

We consider the integration of linear time-dependent PDEs of the form

$$\rho = \rho(R, P, t), \quad \partial_t \rho = \mathcal{L}\rho, \tag{1}$$

where $\rho := \mathcal{R}^{n_{\dim}} \times \mathcal{R}^{n_{\dim}} \times \mathcal{R} \to \mathcal{R}^M$ is a function to be propagated. We assume the differential operator $\mathcal{L}$ has a purely imaginary spectrum. This setting includes the classical and the quantum classical Liouville equations as well as the Schrödinger equation. For the sake of simplicity, however, we will concentrate on the classical Liouville equation

$$\begin{aligned} \partial_t \rho(R, P, t) &= -M^{-1} P^T \nabla_R \rho(R, P, t) + (\nabla_R V(R))^T \nabla_P \rho(R, P, t), \\ \rho(R, P, 0) &= \rho_0 \quad \forall\, R, P, \end{aligned} \tag{2}$$

where the phase space consists of location R and impulse P, M is a symmetric positive definite matrix of masses, $V(R)$ is the potential energy function, and $\rho : \mathcal{R}^{n_{\dim}} \times \mathcal{R}^{n_{\dim}} \times \mathcal{R} \to \mathcal{R}$ is the Wigner quasi density.

As it has been already mentioned, a number of approaches can be applied to integrate the partial differential eq. (2). But most of the existing particle methods for the CLE, both with Dirac and Gauss functions as a basis, share the lack of theoretically justified adaptivity in time and space. To construct a fully adaptive method we employ the Rothe method[31] of semidiscretization in time, which leaves us with a stationary PDE to be solved in each time step. Spatial adaptivity can then be exploited for robust and efficient solution of these stationary problems.

Because the spectrum of the differential operator $\mathcal{L}$ is purely imaginary, Gauss methods are suited for time discretization. We choose the most simple scheme, the well-known implicit trapezoidal rule

$$\left(I - \frac{\tau}{2}\mathcal{L}\right)\bar{\rho}(t + \tau) = \left(I + \frac{\tau}{2}\mathcal{L}\right)\rho(t), \tag{3}$$

where $\bar{\rho}$ is the value obtained by the time-discrete evolution, starting from the (exactly available) initial value ρ. Note that all

Gauss integrators conserve first integrals, which implies conservation of volume and energy for the CLE setting.

For adaptivity in time, we need three essential ingredients: an error estimator, a step size selection scheme, and a desired tolerance. We briefly recollect the standard methodology from integration of ODEs (cf. ref. 32).

Error Estimator

Denoting the exact evolution by Φ, we estimate the unknown error

$$\bar{\varepsilon}_t := \|\bar{\rho}(t+\tau) - \Phi_\tau \rho(t)\| \tag{4}$$

by the difference between the trapezoidal rule and some easily computable comparison propagator $\hat{\Psi}_\tau$ of lower order, for example, the explicit Euler method $\hat{\Psi}_\tau = I + \tau\mathcal{L}$:

$$[\bar{\varepsilon}_t] := \|\bar{\rho}(t+\tau) - \hat{\Psi}_\tau \rho(t)\|. \tag{5}$$

The step is accepted if $[\bar{\varepsilon}_t]$ is sufficiently small, i.e., $[\bar{\varepsilon}_t] \leq \mathrm{TOL}_t$, where TOL_t is a user-prescribed accuracy requirement. Otherwise, we reduce the step size and repeat the step. Note that $[\bar{\varepsilon}_t]$ necessarily estimates the error $\hat{\varepsilon}_t$ of the less accurate comparison propagator $\hat{\Psi}_\tau$ instead of the computationally unavailable error of the trapezoidal rule.

A tempting idea would be to use efficient explicit second-order particle propagators, which have been developed under the LHA and IPA assumptions,[15,22] as $\hat{\Psi}_\tau$. However, in the case of nonharmonic potentials and Gaussians with nonvanishing width, these propagators are of convergence order zero. Although being a reasonable approximation to the exact evolution Φ_τ, such propagators provide worse error estimates than the explicit Euler scheme, which is of convergence order one.

Step Size Selection Scheme

We assume the comparison propagator $\hat{\Psi}_t$ is of order one, such that

$$[\bar{\varepsilon}_t] \doteq \hat{\varepsilon}_t \doteq C\tau^2 \tag{6}$$

holds locally for some slowly varying constant C. Substituting $[\bar{\varepsilon}_t]$ for $\bar{\varepsilon}_t$ and aiming at an error of $\sigma\mathrm{TOL}_t$ with some safety factor $\sigma < 1$, we obtain an optimal step size

$$\tau_{\mathrm{opt}} = \sqrt{\frac{\sigma \mathrm{TOL}_t}{[\bar{\varepsilon}_t]}}\,\tau, \tag{7}$$

which is used for the next step or recomputing the current time step, respectively.

Adaptive Phase Space Discretization

For approximating the distributions to be propagated, we use a linear combination

$$\rho(t) = \sum_{n=1}^{N} y_n(t)\, g(\bar{R}_n(t),\, \bar{P}_n(t),\, \bar{G}_n(t)) \tag{8}$$

of particles g positioned at points $(\bar{R}_n,\, \bar{P}_n)(t)$ in phase space, scaled by the amplitudes $y_n(t)$. Additionally, the shape of the particles is allowed to depend on a set of shape parameters $\bar{G}_n(t)$.

For the CLE, we use Gaussian phase space packets (GPPs) defined as

$$g(\bar{R},\, \bar{P},\, \bar{G})(R,\, P) := \exp\left[-\begin{pmatrix} R-\bar{R} \\ P-\bar{P} \end{pmatrix}^T \bar{G} \begin{pmatrix} R-\bar{R} \\ P-\bar{P} \end{pmatrix} \right], \tag{9}$$

where the shape matrix $\bar{G} \in \mathcal{R}^{4n^2_{\mathrm{dim}}}$ is symmetric positive definite.

To propagate an initial Wigner density by means of the particle discretization given above, two tasks have to be tackled. First, an initial GPP approximation of a given Wigner density has to be computed, and second, in every time step a new GPP approximation to the exact propagation of the current GPP approximation must be found.

Initial GPP Approximation

A method to approximate a given Wigner density ρ_0 by a linear combination $\rho(0)$ of few GPPs has been recently proposed by Horenko, Schmidt, and Schütte.[33] Because similar techniques are developed in the next section for spatial adaptivity, we sketch the method here for convenience.

The aim is to achieve a sufficiently small spatial approximation error

$$\|\rho(0) - \rho_0\|_{\mathcal{L}_1} \leq \mathrm{TOL}_x$$

subject to $\bar{G}_n$ symmetric positive definite with some number N of GPPs to be determined as small as possible. To make the task computationally tractable, we substitute the $\mathcal{L}_1$-norm by a discrete Monte Carlo sampling at K_N points and simplify the positive definiteness constraint to fixing $\bar{G}_n = \lambda I$ with some $\lambda > 0$, obtaining the requirement

$$\|\rho(0) - \rho_0\|^2_{\{R_k,P_k\}} := \sum_{k=1}^{K_N} \omega_k |\rho_0(R_k, P_k) - \rho(R_k, P_k, 0)|^2 \leq \mathrm{TOL}_x^2. \tag{10}$$

For (10) not to be underdetermined, we have to select at least as many sample points as the representation (8) has degrees of freedom, and hence, require $K_N \geq \#\mathrm{dof}(N)$. For fixed N, the approximation error can be minimized by a Gauss–Newton method. Initial centers $(\bar{R}_n,\, \bar{P}_n)$ of the GPPs and sampling points $(R_k,\, P_k)$ are obtained by a Monte Carlo sampling according to the absolute value of the quasi-probability density ρ_0. For computational feasibility we restrict the sampling to the regions in phase space where $|\rho_0|$ exceeds a certain threshold. For simplicity, we include the GPPs centers into the set of sampling points by setting $(R_k,\, P_k) = (\bar{R}_n,\, \bar{P}_n)$ for $1 \leq k \leq N$ and generate at least $\#\mathrm{dof}(N) - N$ more sampling points by the same Monte Carlo

process. In concordance with the probabilistic density of the sampling points, the weights have to be chosen as

$$\omega_k := \frac{\|\rho_0\|_{\mathscr{L}_1}}{K_N \rho_0(R_k, P_k)}.$$

If the local optimum computed by the Gauss–Newton method does not satisfy (10), N is increased and additional GPP's are created by Monte Carlo sampling. The process is then repeated until the accuracy requirement is fulfilled.

Propagation of GPP Approximations

Numerical Propagation

The propagation of the Wigner density ρ by the implicit trapezoidal rule (3) poses an approximation problem similar to that encountered in the previous section, namely to find a new density $\bar{\rho}(t + \tau)$ representable by GPPs such that

$$\varepsilon_x = \left\| \left(I - \frac{\tau}{2} \mathscr{L}_c \right) \bar{\rho}(t + \tau) - \left(I + \frac{\tau}{2} \mathscr{L}_c \right) \rho(t) \right\|_{\mathscr{L}_1} \leq \mathrm{TOL}_x. \quad (11)$$

Here, TOL_x is a tolerance that now has to be matched with the user-prescribed accuracy requirement TOL_t—see below. By sampling at K_N points (R_i, P_i), we again reduce (11) to a computationally tractable least squares problem

$$[\varepsilon_x] = \left\| \left(I - \frac{\tau}{2} \mathscr{L}_c \right) \rho(t + \tau) - \left(I + \frac{\tau}{2} \mathscr{L}_c \right) \rho(t) \right\|_{\{R_i, P_i\}} \leq \mathrm{TOL}_x. \quad (12)$$

Depending on how the degrees of freedom to be fitted by the least-squares procedure are selected, either only the amplitudes, or both amplitudes and phase space positions of the GPPs' centers, we arrive at a linear or nonlinear least squares problem. In the linear version, the system has K_N equations and $\#\mathrm{dof}(N) = N$ unknowns y_n defining $\rho(t + \tau)$, and can be solved by a single QR decomposition of the influence matrix

$$A = \frac{\partial}{\partial y} \left[\left(I - \frac{\tau}{2} \mathscr{L}_c \right) \rho(t + \tau) \right].$$

In the nonlinear variant, the system has K_N equations and $\#\mathrm{dof}(N) = (1 + 2n_{\mathrm{dim}})N$ degrees of freedom y_n, $\bar{R}_n$, and $\bar{P}_n$ defining $\rho(t + \tau)$. Due to the better approximation capability offered by also adjusting the GPPs' centers, the number of GPPs necessary to satisfy the accuracy requirement (12) can be expected to be considerably smaller than for the linear approach. However, this does not necessarily translate into fewer degrees of freedom, or fewer sample points. For solving the nonlinear least-squares problem, a Gauss–Newton method should be used, which may require multiple QR decompositions of the influence matrix

$$A = \frac{\partial}{\partial (y, \bar{R}, \bar{P})} \left[\left(I - \frac{\tau}{2} \mathscr{L}_c \right) \rho(t + \tau) \right].$$

Whether this is compensated by the better approximation capability is not clear *a priori.*

Choice of GPP Collection

There are several possibilities to choose the GPP collection used to represent $\rho(t + \tau)$ in the beginning of each time step. Selecting GPPs in unsuitable regions of the phase space will prevent the linear least-squares approach from meeting the accuracy requirement (12), thus triggering the discretization refinement developed below. For the nonlinear least-squares approach, it will increase the number of Gauss–Newton steps, and hence, decrease the computational efficiency. A sufficiently good initial guess for the solution of the least-squares problem is therefore necessary for computational efficiency in both variants.

The Monte Carlo generation of the centers $(\bar{R}_n, \bar{P}_n)$ anew for every least-squares problem is possible, but too costly. Fortunately, it can be omitted here due to the continuity of the evolution: because for sufficiently small time steps τ the new density $\rho(t + \tau)$ is close to the old one, we can expect that the Gauss–Newton method starting at the old density $\rho(t)$ converges quickly towards the closest local minimum of the approximation error (12). Similarly, the linear-least squares solution can be expected to be sufficiently accurate. By cheaply computable approximations to the evolution, even better initial guesses can be provided, thus enabling larger time steps and reducing the need for additional GPPs. Two such *predictors* are presented below.

Choice of Sample Points

Another question that has to be addressed is the choice of sample points (R_i, P_i), $i = 1, \ldots, K_N$. For the least-squares problem (12) not to be underdetermined, we require at least $K_N \geq \#\mathrm{dof}(N)$ sample points, preferably distributed in accordance with the quasi-probability density $\rho(t)$. K_N should be significantly larger than $\#\mathrm{dof}(N)$ to improve the robustness of the least-squares approximation and to provide a local error estimator for spatial adaptivity—see below. Because performing a Monte Carlo sampling at every time step is prohibitively expensive, we suggest to select the sampling points according to the following scheme: For the first step at $t = 0$ we take the sample points from the initial GPP approximation outlined earlier. For subsequent steps, we suggest to take again the centers of the GPP's, i.e. $[R_i(t + \tau), P_i(t + \tau)) = (\bar{R}_i(t + \tau), \bar{P}_i(t + \tau)]$, $i = 1, \ldots, N$, and additionally the remaining sampling points from the previous step propagated independently of each other in time along classical trajectories, i.e., $[R_i(t + \tau), P_i(t + \tau)] = \Phi_\tau(R_i(t), P_i(t))$, $i = N + 1, \ldots, K_N$.

Spatial Adaptivity

It may happen that the number N of GPPs chosen to fit the initial state $\rho(0)$ becomes inadequate during the propagation, for three different reasons. (a) A more complicated distribution $\rho(t)$ arises later in time, such that more GPPs are needed to represent the

distribution $\bar{\rho}(t)$ with the required accuracy. (b) Two or more GPPs can come close to each other, such that the least squares problem (12) becomes ill conditioned. (c) The distribution may develop a simpler structure, such that it is advisable to reduce the number of GPP's for computational efficiency. The first situation requires the generation (*spawning*) of new GPPs, whereas the latter ones require the removal (*pruning*) of existing GPPs.

Let us first consider the case that the number of GPPs is too small, such that the accuracy requirement (12) cannot be satisfied. In this case, as few as possible additional GPP's have to be created to reduce the approximation error sufficiently. Fortunately, the local residuals

$$\varepsilon_k = \left| \left(I - \frac{\tau}{2} \mathcal{L} \right) \rho(R_k, P_k, t + \tau) - \left(I + \frac{\tau}{2} \mathcal{L} \right) \rho(R_k, P_k, t) \right| \tag{13}$$

provide a useful local error indicator suitable for extending the particle set. A similar error indicator has been proposed by Iske and Levesley[34] in the context of scattered data approximation. The following scheme is intended to insert the new particles at positions in phase-space, where the approximation error is largest, and hence, to improve the approximation at a small cost.

Assume the sample points (R_k, P_k), $k = N + 1, \ldots, K_N$ (which are not also centers of existing GPPs), are sorted descendingly by their local residual $\omega_k \varepsilon_k$. Let $j > N$ be minimal such that

$$\sum_{k=j+1}^{K_N} \omega_k \varepsilon_k \leq \mathrm{TOL}_x \tag{14}$$

holds, or $j = K_N$ if (14) cannot be satisfied. We then suggest to substitute (or "upgrade") the sample points $N + 1, \ldots, j$ by newly created GPPs with centers (R_k, P_k), $k = N + 1, \ldots, j$, amplitude zero, and shape matrix λI, and create at least $2n_{\mathrm{dim}}(K_N - j)$ new sample points in the vicinity of the newly created GPPs by some Monte Carlo method. N and K_N should be increased accordingly to j and $K_N + 2n_{\mathrm{dim}}(K_N - j)$, respectively. With the enlarged particle set at hand, the least-squares problem is solved again to meet the requirement (12). If necessary, the adaptive refinement is repeated until finally (12) is met. Related greedy algorithms for spatial adaptivity in different contexts have been proposed by Schaback et al.[35,36]

Let us now turn to the case that the least-squares problem (12) becomes ill conditioned due to similarly shaped GPPs being too close to each other. Sawada et al.[7] suggest to drop an arbitrary Gaussian and do a refitting of the remaining ones whenever one eigenvalue of the overlap matrix becomes small. Although this criterion is reported to work, it neither takes the approximation error into account nor does it indicate which Gaussian to drop or how small an eigenvalue must become. Wan et al.[26] suggest to remove Gaussians with an amplitude below 10^{-12} and to collapse any two Gaussians that are too close to each other. Although this can indeed cure the numerical stability problems, no indication is given how close two Gaussians must be or how the cutoff value of the amplitude was chosen.

As a pruning method oriented at the numerical stability and the approximation error, we propose to use a column permutation strategy[37] for the QR decomposition together with a numerical rank decision based on the subcondition number[38] to identify and remove exactly those columns and their associated GPPs that make (12) numerically singular. Moreover, a careful examination of the least-squares residual enables the identification of even more GPPs, which are not necessary to obtain the requested accuracy, and thus can be removed. Pruning of the GPP collection should be realized by "downgrading" unnecessary Gaussians to sample points.

To be more precise, in the linear least squares setting, assume the columns of A and correspondingly the rows of x of the linear least squares problem, $\|Ax - b\| = \min$, have been sorted such that for the QR decomposition $A = QR$ the relations

$$|R_{ii}| \geq |R_{i+1,i+1}\| \quad \text{for } i = 1, \ldots, \#\mathrm{dof}(N) - 1$$

hold. Construct a partition

$$R = \begin{pmatrix} R_1 & S_1 & S_2 \\ & R_2 & S_3 \\ & & R_3 \\ & & 0 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad Q^T b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}, \tag{15}$$

such that the following conditions are satisfied:

$$\max_i |(R_3)_{ii}| \leq \frac{1}{\kappa} |(R_1)_{11}| < \min_i |(R_2)_{ii}| \sqrt{\|b_2\|^2 + \|b_3\|^2 + \|b_4\|^2}$$
$$\leq \sigma \mathrm{TOL}_x$$

for κ being the maximal accepted least-squares condition (somewhere around $\kappa = 10^8$) and $0 < \sigma < 1$ some safety factor that can be adjusted to balance pruning and spawning. A default value of $\sigma = 0.9$ is suggested. Note that $\|b_4\|$ is the smallest possible approximation error that can be achieved at all with the present collection of Gaussians, and similarly $\sqrt{\|b_3\|^2 + \|b_4\|^2}$ is the minimal error that can be obtained in a numerically stable way.

The columns of A and GPPs corresponding to the degrees of freedom in x_3 can be removed on the observation that they are numerically linearly dependent on the columns corresponding to x_1 and x_2, and, hence, are redundant.

Furthermore, the degrees of freedom in x_2 contribute least to the approximation capability of the remaining GPP collection. Sacrificing some accuracy while still satisfying the accuracy requirement (12) allows to improve the computational efficiency.

In case no such partition can be found, i.e., $\sigma \mathrm{TOL}_x < \sqrt{\|b_3\|^2 + \|b_4\|^2} \leq \mathrm{TOL}_x$, just the numerically linearly dependent degrees of freedom are pruned. If the accuracy requirement (12) cannot be fulfilled at all, i.e., $\sqrt{\|b_3\|^2 + \|b_4\|^2} > \mathrm{TOL}_x$, the spawning procedure described above has to be performed.

In the setting of the nonlinear least squares fitting, the correspondence of columns in A to GPPs is no longer one to one, such that the pruning procedure described above has to be modified. Numerical stability even in the case of linearly dependent columns can be maintained by setting $x_3 = 0$ without removing the

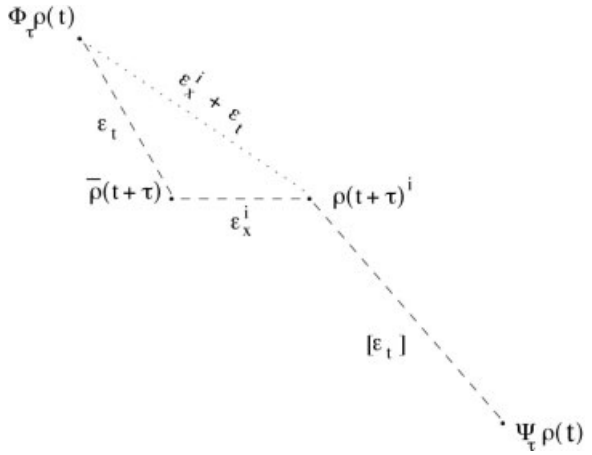

Figure 1. Time and spatial discretization errors in the Rothe method for the exact evolution $\Phi_\tau\rho(t)$. The time step size control based on the difference $[\varepsilon_t]^i$ between explicit Euler $\Psi_\tau\rho(t)$ and discretized trapezoidal rule $\rho(t+\tau)^i$ requires the spatial discretization error ε_x^i to be comparable to the time error ε_t of the exact trapezoidal rule $\bar{\rho}(t+\tau)$.

corresponding degrees of freedom. We suggest to perform the last Gauss–Newton step, when the GPP's centers are already close to the solution, in a reduced fashion by fitting only the amplitudes. In this way, the pruning scheme developed for the linear least-squares case can be transferred to the nonlinear case as well.

Accuracy Matching

Unfortunately, the time error estimator $[\bar{\varepsilon}_t] = \|\bar{\rho}(t+\tau) - \hat{\Psi}_\tau\rho(t)\|$ from earlier is still computationally unavailable. As shown in Figure 1, its canonical substitute $[\varepsilon_t] := \|\rho(t+\tau) - \hat{\Psi}_\tau\rho(t)\|_{\{R_i,P_i\}}$ depends on the spatial discretization error ε_x, which should not destroy the overall quality of the error estimate. In view of $[\bar{\varepsilon}_t] \doteq C\tau^2$ and $\varepsilon_t \doteq C\tau^3$, and in order not to destroy the second-order convergence of the trapezoidal rule, we aim at $\varepsilon_x \doteq \varepsilon_t$ and, hence, impose the accuracy matching $\mathrm{TOL}_x = \tau\mathrm{TOL}_t$.

Asymptotic Conservation Properties

The adaptive refinement described above recovers the conservation of energy and volume featured by the exact trapezoidal rule (3) asymptotically for $\mathrm{TOL}_x \to 0$. Assuming the absolute value of the potential V to be bounded by some polynomial $p > 0$ of degree at least 2, the energy

$$\langle E, \rho\rangle = \int \left(V(R) + \frac{1}{2} PM^{-1}P^T\right)\rho dP dR$$

is a continuous linear functional on the weighted L_1-space Y defined by the norm $\|y\|_Y = \|py\|_{L_1}$. Let p, in turn, be bounded by some constant $C < \infty$ on the bounded region of the phase space where the GPPs are located. Then any sensible particle approximation $\rho(t+\tau)$ of $\bar{\rho}(t+\tau)$ that satisfies (12) will also fulfill

$$\left\|\left(I - \frac{\tau}{2}\mathcal{L}\right)\rho(t+\tau) - \left(I + \frac{\tau}{2}\mathcal{L}\right)\rho(t)\right\|_Y \le C\mathrm{TOL}_x.$$

Because the exact trapezoidal rule conserves quadratic first integrals, the energy error of the approximate solution $\rho(t+\tau)$ can thus be bounded by

$$\begin{aligned}
\varepsilon_E(t) &:= |\langle E, \rho(t+\tau) - \Phi_\tau\rho(t)\rangle| \\
&\le |\langle E, \rho(t+\tau) - \bar{\rho}(t+\tau)\rangle| + \underbrace{|\langle E, \bar{\rho}(t+\tau) - \Phi_\tau\rho(t)\rangle|}_{=0} \\
&\le \|E\|\,\|\rho(t+\tau) - \bar{\rho}(t+\tau)\|_Y \\
&\le \|E\|\left\|\left(I - \frac{\tau}{2}\mathcal{L}\right)^{-1}\right\| \\
&\qquad \times \left\|\left(I - \frac{\tau}{2}\mathcal{L}\right)(\rho(t+\tau) - \bar{\rho}(t+\tau))\right\|_Y \\
&\le \|E\|\left\|\left(I - \frac{\tau}{2}\mathcal{L}\right)\rho(t+\tau) - \left(I + \frac{\tau}{2}\mathcal{L}\right)\rho(t)\right\|_Y \\
&\le \|E\| C\mathrm{TOL}_x.
\end{aligned}$$

Here, we have used that the differential operator $\mathcal{L}$ has an unbounded, purely imaginary spectrum, which implies $\|(I - (\tau/2)\mathcal{L})^{-1}\| = 1$.

Analogously, asymptotic conservation of volume can be shown, or, for different evolutions, conservation of arbitrary quadratic first integrals.

Note that this result does not guarantee long term conservation of energy, as has been established for the method of lines, i.e., semidiscretization in space, by Hairer, Lubich, and Wanner.[39]

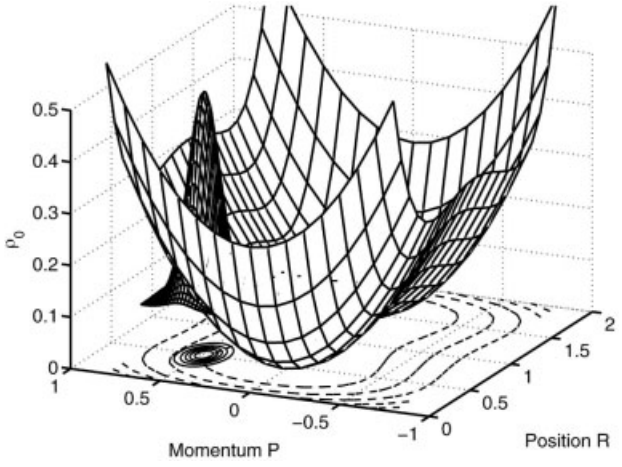

Figure 2. Potential energy surface $V(R, P)$ and initial Wigner density in phase-space representation.

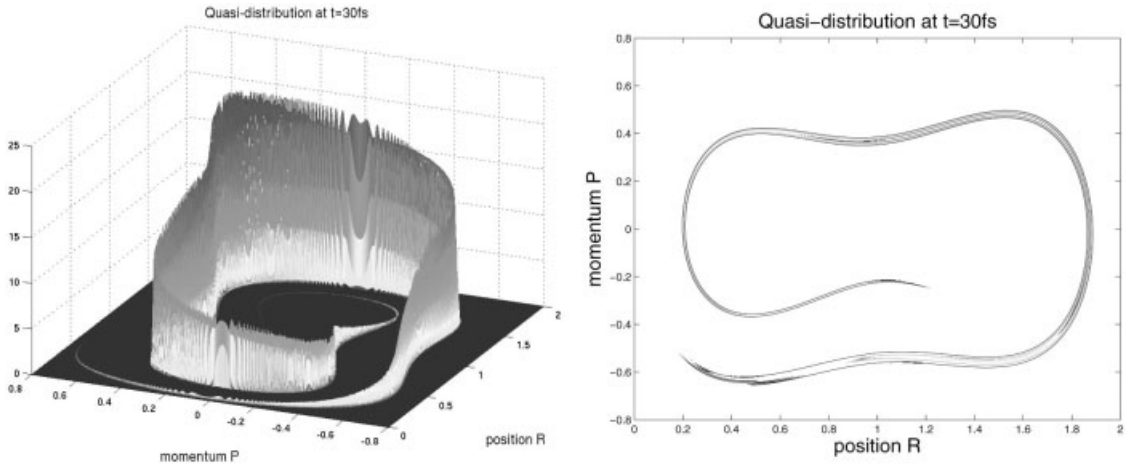

Figure 3. Quasi-distribution at $t = 30$ fs, evaluated on a 700×800 grid. Contour lines are drawn at levels 1, 5, 10, and 15.

Predictors

The simplest choice of the starting point for the Gauss–Newton method is, of course, the current GPP collection $\rho(t)$. However, because $\rho(t + \tau) - \rho(t) = \mathbb{O}(\tau)$, the time step τ is limited by the requirement that the initial guess should be sufficiently good such that the local Gauss–Newton iteration converges quickly and reliably to the nearest local solution. Similarly, if only the amplitudes y_n are fitted by a linear least-squares approach, a good initial guess yields a particle set that is well suited to represent the solution. Thus, the accuracy requirement (12) can be satisfied with fewer GPPs.

For these reasons, the employment of a cheaply computable predictor providing a better initial guess can be expected to improve the performance of the propagation considerably, by allowing larger time steps (in the Gauss–Newton case) and by decreasing the number of necessary GPPs (in the linear least-squares case).

For the CLE considered here, we suggest using an explicit symplectic modified Leap-Frog propagator Ψ recently proposed by Horenko, Schmidt, and Schütte[33] as predictor.

In the simple case of both LHA and IPA holding, the Gauss particles in the ensemble can be propagated independently with evolution equations for the parameters R_n, P_n, and G_n derived from (2):

$$\partial_t \bar{R}_n = M^{-1}\bar{P}_n \tag{16}$$

$$\partial_t \bar{P}_n = -\nabla_R V(\bar{R}_n) \tag{17}$$

$$\partial_t \bar{G}_n = C(\bar{R}_n)\bar{G}_n + \bar{G}_n C^T(\bar{R}_n), \tag{18}$$

where

$$C(\bar{R}_n) = \begin{pmatrix} 0 & \nabla_R^2(\bar{R}_n) \\ -M^{-1} & 0 \end{pmatrix}.$$

For sufficiently narrow GPPs, the LHA holds at least approximately even for nonharmonic potentials, such that the predictor solution $\Psi_\tau \rho(t)$ can be expected to provide a good approximation of the exact solution $\Phi_\tau \rho(t)$.

This scheme works well for low to medium dimensional problems, but requires the second derivative of the potential to be evaluated. This evaluation can become expensive for higher dimensional problems. For such cases, the adaptive semidiscretization in time provides the freedom to use *any* sensible predictor that promises an efficient representation of $\rho(t + \tau)$. Even if the predictor is not physically justified, it will not impair the quality of the solution, but may affect the efficiency of the algorithm, for example, in the numerical example of the next section we used a simple heuristic predictor that stretches particles along the flow direction to represent the highly anisotropic density well. An even simpler choice, of course, is to use "frozen Gaussians," where only the centers of the particles are propagated by the predictor, but their shape remains fixed.

Algorithm. Overall, we end up with the following algorithm.

Initial phase-space distribution approximation (GPP decomposition):
- $N := \mathrm{TOL}_x^{-1/2}$
- direct Monte Carlo generation of $(\bar{R}_n, \bar{P}_n)$, $n = 1, \ldots, N$
- $K_N := 2Nn_{\mathrm{dim}}$
- direct Monte Carlo generation of (R_k, P_k), $k = 1, \ldots, K_N$
- solve nonlinear approximation problem (10) for y_n, $\bar{R}_n$, $\bar{P}_n$
- **while** $\|\rho_0 - \Sigma_{n=1}^{N} y_n g(\bar{R}_n, \bar{P}_n, \lambda I)\|_{\{R_k, P_k\}} > \mathrm{TOL}_x$:
  - $N := 1.1N$
  - $K_N := 2Nn_{\mathrm{dim}}$
  - direct Monte Carlo generation of new GPPs and sample points
  - solve nonlinear approximation problem (10) for y_n, $(\bar{R}_n, \bar{P}_n)$

Numerical propagation of GPP distribution:

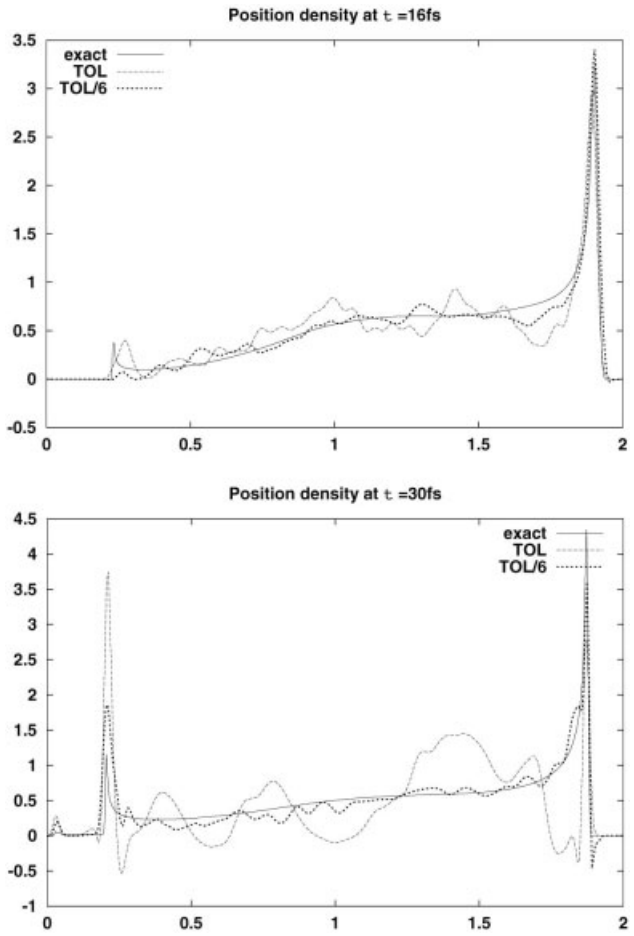

Figure 4. Snapshots of position space representation of the Wigner density evolution at times 16 fs and 30 fs as obtained from the special purpose (solid line) and adaptive methods (at different tolerances).

while $t < T$:
  compute predictor $\Psi_\tau \rho(t)$
  solve (non)linear approximation problem (12) for y_n, $(\bar{R}_n, \bar{P}_n)$
  remove unnecessary GPP's using the partition (15)
  while (12) not satisfied:
    $N := j$ with j from (14)
    $K_N := 2Nn_{\text{dim}}$
    local generation of new GPPs and sample points
    solve (non)linear approximation problem (12) for y_n, $(\bar{R}_n, \bar{P}_n)$
  compute error estimator $[\varepsilon]$ from (5)
  $t := t + \tau$
  $\tau := \sqrt{\sigma \text{TOL}_t/[\varepsilon]}\ \tau$

Numerical Example

As an example for the application of the proposed TRAIL scheme we consider a one-dimensional model of the Gauss-shaped density of unitary mass initially centered at 0.5 a.u. of length with width parameter 0.5 and momentum 0.48 a.u. in a double-well potential (Fig. 2). This model can, for example, qualitatively describe the proton-transfer process in proteins or liquids.[41,42] In this model covalent bonds are described as coupled harmonic oscillators, which, in adiabatical representation, produce a double-well potential for a lower adiabatic state. Assuming tunneling and nonadiabatical coupling to be negligible, the dynamics will be governed by classical transport along the adiabatical surface. Reference solutions for such systems can be computed by standard Leap-Frog propagation of classical trajectories, where the values of the quasi-density are kept constant at the centers of the particles.

Despite its apparent simplicity, this example develops quasi-densities with highly local features (see Fig. 3) and is therefore quite challenging for any numerical scheme.

In the numerical examples we found the linear-least squares approach, where only the amplitudes y_n are fitted, more robust and efficient than the full nonlinear least-squares method. The solutions shown below were therefore obtained with the linear approach. A simple heuristic predictor has been used, which stretches particles along the flow direction to represent the highly anisotropic density well.

The numerical implicit trapezoidal rule (3) at different tolerances is compared with an almost exact special purpose method exploiting the flow conservation property of this particular exam-

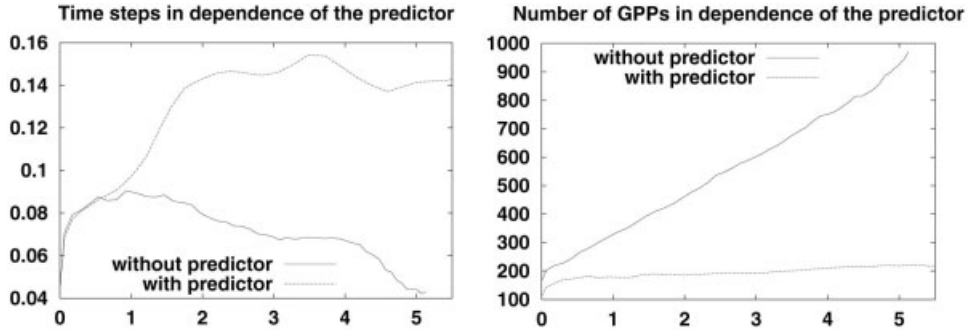

Figure 5. Time steps and number of GPPs in dependence of the predictor. In case no predictor is employed to move the particles with the flow, the adaptive spatial discretization is able to follow the flow just by creating new particles. However, the performance gain that can be obtained by using a suitable predictor as outlined in the text is dramatic.

ple. The initial Wigner density is decomposed into an ensemble of 100 GPPs with global error of 5%. Figure 4 shows position representations of the density evolution on a time span of 30 fs as obtained from the special purpose (solid line) and adaptive (dashed lines) methods.

Another numerical example shows the efficiency of the space-adaptive algorithm in dependence of the predictor (Fig. 5). Not surprisingly, the pictures clearly show that an effective predictor is a key ingredient for an adaptive particle method.

Conclusion

A particle method with full temporal and spatial error estimation and adaptivity has been developed for Liouville-type equations. Due to adaptivity, the method is able to reliably represent sophisticated spatiotemporal details of the dynamics under consideration.

The computational cost of the method is increasing with the geometric complexity of obtained solutions.

Closed dynamical systems develop more and finer details as the propagation time increases. This property of Hamiltonian dynamics constricts virtually all numerical schemes to comparably low spatial dimension and short propagation times.

For open dynamical systems, for example, with inclusion of stochastic influence, the solution will typically be much smoother and therefore much more appropriate for GPP approximation with wide GPPs even in higher dimension.

Acknowledgment

The authors would like to thank C. Schütte for helpful suggestions, and W. Huisinga and B. Schmidt for careful reading of the manuscript.

References

1. Zenger, C. In Parallel Algorithms for Partial Differential Equations; Proc 6th GAMM-Seminar, Kiel 1990. Notes Numer Fluid Mech 1991, 31, 241.
2. Griebel, M.; Zumbusch, G. In Proceedings of the 7th International Conference on Hyperbolic Problems, Theory, Numerics, Applications; Birkhäuser: Boston, MA, 1998.
3. Neunzert, H.; Klar, A.; Struckmeier, J. In ICIAM 95: Proceedings of the Third International Congress on Industrial and Applied Mathematics; Kirchgaessner, K., Ed.; Hamburg, Germany, 1995.
4. Griebel, M.; Schweitzer, M. A., Eds. Meshfree Methods for Partial Differential Equations. Lecture Notes in Computational Science and Engineering 26; Springer: Berlin, 2002.
5. Heller, E. J. Quantum corrections to classical photodissociation models. J Chem Phys 1978, 68, 2066.
6. Heller, E. J. Acc Chem Res 1981, 14, 368.
7. Sawada, S.-I.; Heather, R. J Chem Phys 1985, 83, 3009.
8. Sawada, S.-I.; Metiu, H. J Chem Phys 1986, 84, 6293.
9. Kapral, R.; Ciccotti, G. J Chem Phys 1999, 110, 8919.
10. Schütte, Ch. Konrad-Zuse-Center, Preprint SC-99-10, 1999. available through http://www.zib.de/bib.
11. Nielsen, S.; Kapral, R.; Ciccotti, G. J Chem Phys 2000, 112, 6543.
12. Hillery, M.; O'Connell, R. F.; Scully, M. O.; Wigner, E. P. Phys Rep 1984, 106, 121.
13. Ma, J.; Hsu, D.; Straub, J. E. J Chem Phys 1993, 99, 4024.
14. Schleich, W. P. Quantum Optics in Phase Space; Wiley-VCH: Berlin, 2001.
15. Horenko, I.; Schmidt, B.; Schütte, Ch. J Chem Phys 2002, 117, 4643.
16. Wigner, E. P. Phys Rev 1932, 40, 749.
17. Martinez, T. J.; Ben-Nun, M. J Chem Phys 1996, 104, 2847.
18. Martinez, T. J.; Ben-Nun, M. J Phys Chem 1996, 100, 7884.
19. Ben-Nun, M.; Martinez, T. J. J Chem Phys 1998, 108, 7244.
20. Horenko, I.; Salzmann, Ch.; Schmidt, B.; Schütte, Ch. J Chem Phys 2002, 117, 11075.
21. Donoso, A.; Martens, C. C. Phys Rev Lett 2001, 87, 223202.
22. Heller, E. J. J Chem Phys 1975, 62, 1544.
23. Brown, R. C.; Heller, E. J. J Chem Phys 1981, 75, 186.
24. Drolshagen, G.; Heller, E. J. J Chem Phys 1981, 75, 186.
25. Drolshagen, G.; Heller, E. J. J Chem Phys 1981, 75, 186.
26. Wan, C.-C.; Schofield, J. J Chem Phys 1999, 112, 4447.
27. Wan, C.-C.; Schofield, J. J Chem Phys 2000, 113, 7047.
28. Wan, C.-C.; Schofield, J. J Chem Phys 2002, 116, 494.
29. Walkup, R. J Chem Phys 1991, 95, 6440.
30. Prezhdo, O.; Kisil, V. V. Phys Rev A 1997, 56, 162.
31. Bornemann, F. IMPACT Comput Sci Eng 1990, 2, 279.
32. Deuflhard, P.; Bornemann, F. Numerische Mathematik II: Integration gewöhnlicher Differentialgleichungen; Walter de Gruyter: Berlin, 1994.
33. Horenko, I.; Schmidt, B.; Schütte, Ch. J Chem Phys 2002, 117, 4643.
34. Iske, A.; Levesley, J. Technical Report, University of Leicester, 2002.
35. Schaback, R.; Wendland, H. Numer Algorithms 2000, 24, 239.
36. Schaback, R.; Hon, Y. C.; Zhoud, X. Preprint, 2001.
37. Businger, P.; Golub, G. H. Linear least squares solutions by Householder transformations. Numer Math 1965, 7, 269.
38. Deuflhard, P.; Sautter, W. Lin Algebra Appl 1980, 29, 91.
39. Hairer, E.; Lubich, Ch.; Wanner, G. Geometric Numerical Integration; Springer: Berlin, 2002.
40. Strang, G. SIAM J Numer Anal 1968, 5, 506.
41. Schuster, P.; Mikenda, W. Hydrogen Bond Research; Springer: Berlin, 1999.
42. Hynes, J. T.; Tran-Thi, T. H.; Granucci, G. J Photochem Photobio A 2002, 154, 3.