Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Physics Communications 169 (2005) 339–342
Computer Physics Communications
www.elsevier.com/locate/cpc

Linear prediction of force time series to accelerate molecular dynamics simulations

Branislav Brutovsky [a], Gerald R. Kneller [b,c,*]

[a] *Institute of Physics, P.J. Safarik University, Kosice, Slovakia*
[b] *Laboratoire Léon Brillouin, CEA Saclay, France*
[c] *Centre de Biophysique Molèculaire, CNRS, Orléans, France*

Available online 9 April 2005

Abstract

We have recently proposed a molecular dynamics simulation scheme in which the time-consuming evaluation of non-bonding forces is periodically replaced by linear prediction of the latter from previous values [B. Brutovsky, T. Mülders, G.R. Kneller, J. Chem. Phys. 118 (2003) 6179]. For a simple molecular liquid, consisting of linear molecules with an internal vibrational degree of freedom, the method yields a speedup of up to 7 compared to conventional simulations. The hybrid simulation scheme preserves all essential structural and dynamical quantities. We show here that the linear predictor can be considered as an optimal integrator for finite time steps whose coefficients approach those of a discrete Taylor series if the simulation step tends to zero. The short time dynamics and the structure of the liquid are preserved for much longer periods if linear prediction is used instead of Taylor expansion to predict the non-bonding forces.

PACS: 02.70.Ns; 07.05.Tp

Keywords: Linear prediction; Time series; Molecular dynamics simulations

1. Introduction

Molecular dynamics (MD) [1] simulations are a well established tool for the study of a broad range of phenomena in chemistry, physics, and biology. The accessible time scale depends crucially on the efficiency of the MD simulation scheme. Therefore, since the early days of MD simulation considerable efforts have been made to develop efficient integrators. An example is the multiple-time-step (MTS) method, where slowly varying long-ranged forces are less often calculated than rapidly varying local forces. In a refined MTS method [2] the long-ranged forces are *predicted* on the basis of a Taylor expansion, instead of keeping them constant during the updates of the local forces. In our recent work [3] we presented an MTS scheme where the explicit calculation of long-ranged

* Corresponding author.
E-mail addresses: bru@seneca.science.upjs.sk (B. Brutovsky), kneller@dsm-mail.cea.fr, kneller@cnrs-orleans.fr (G.R. Kneller).

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2005.03.076

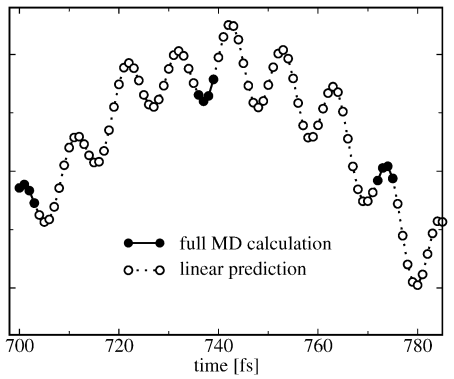

Fig. 1. The "hybrid" simulation scheme in Ref. [3]. Here, 4 explicit MD steps are followed by 32 predicted steps.

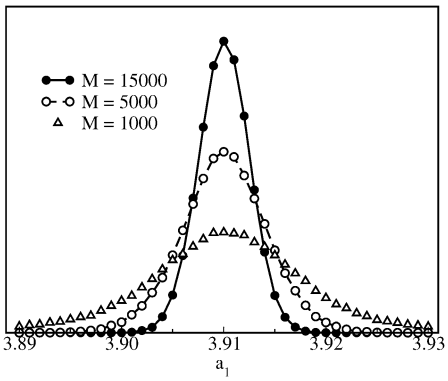

Fig. 2. The typical distribution of the optimal predictor coefficients obtained from Eq. (6), varying the number of observations M. To create the above distributions 100 000 **F**-matrices have been randomly generated for each of the above number of observations.

forces is periodically replaced by linear prediction (LP) (Fig. 1).

2. Theory

In the framework of linear prediction the value of a signal $f(t)$ is estimated from its history through

$$\hat{f}(t) = \sum_{n=1}^{P} a_n f(t - n\Delta t). \tag{1}$$

Here P is the order of the predictor, Δt is a fixed sampling time step, and the coefficients $\{a_n\}$ minimize the time (or ensemble) average of the prediction error $\epsilon(t) := f(t) - \hat{f}(t)$. The fact that MD trajectories can be considered as discrete signals allows easily to build linear estimators for future values of dynamical variables.

To calculate the optimal predictor coefficients $\{a_n\}$ for a force trajectory one computes the M-dimensional vector of observations

$$\mathbf{f} := \left[f_{m_1}(t_1), \ldots, f_{m_M}(t_M) \right]^{\mathrm{T}} \tag{2}$$

and the $M \times P$ matrix $\mathbf{F}$

$$\mathbf{F} := \begin{bmatrix} f_{m_1}(t_1 - \Delta t) & \ldots & f_{m_1}(t_1 - P\Delta t) \\ f_{m_2}(t_2 - \Delta t) & \ldots & f_{m_2}(t_2 - P\Delta t) \\ \vdots & & \vdots \\ f_{m_M}(t_M - \Delta t) & \ldots & f_{m_M}(t_M - P\Delta t) \end{bmatrix} \tag{3}$$

from a short initial MD trajectory. The observation times $\{t_k\}$ are M randomly chosen points on the time axis, and for each time t_k ($k = 1, \ldots, M$) we chose a trajectory $f_{m_k}(t)$. If N is the number of atoms and N_t the number of configurations in the MD trajectory, we have $P\Delta t \leqslant t_k \leqslant (N_t - 1)\Delta t$ and $1 \leqslant m_k \leqslant 3N$. In order to obtain a reliable estimate, one must require that $M \gg P$ (see Fig. 2).

Defining $\mathbf{a} \equiv [a_1, \ldots, a_P]^{\mathrm{T}}$, we try to find $\mathbf{a}$ such that

$$\mathbf{f} \approx \mathbf{F} \cdot \mathbf{a}, \tag{4}$$

where "$\approx$" stands for "as close as possible to" in a least-squares sense. The predictor coefficients are thus determined from the condition

$$\|\mathbf{f} - \mathbf{F} \cdot \mathbf{a}\|^2 = \mathrm{Min}(\mathbf{a}). \tag{5}$$

The unique solution of minimum norm of (5) can be written in the form

$$\mathbf{a}_{\mathrm{opt}} = \mathbf{F}^{+} \cdot \mathbf{f}, \tag{6}$$

where $\mathbf{F}^{+}$ is the generalized inverse of $\mathbf{F}$. We note that $\mathbf{a}_{\mathrm{opt}}$ is the solution of the well-known Gaussian normal equations,

$$(\mathbf{F}^{\mathrm{T}} \cdot \mathbf{F}) \cdot \mathbf{a} = \mathbf{F}^{\mathrm{T}} \cdot \mathbf{f} \tag{7}$$

in case that all columns in $\mathbf{F}$ are linearly independent.

3. Application

The simulated system consists of 1200 flexible oxygen molecules in the liquid phase at a temperature of 73.0 K. The non-bonding intermolecular

forces are modeled by a Lennard–Jones potential ($\epsilon = 0.512167$ a.m.u. $\times$ nm^2/ps^2 and $\sigma = 0.295$ nm), and the internal 0–0 stretching motions by a harmonic force with a force constant corresponding to a vibration frequency $\nu = 47.5$ THz. The simulation time step in all simulations is $\Delta t = 1$ fs.

In order to minimize the jumps of the total intermolecular force on each atom due to partner atoms which cross the cut-off boundary, we use an appropriate smoothing function

$$\phi(r) = \frac{(R_C - r)^2(R_C + 2r - 3R_S)}{(R_C - R_S)^3}, \quad R_S < r < R_C, \tag{8}$$

with $R_C = 0.84$ nm and $R_S = 0.8$ nm. The method is simpler as the one described in [3] and yields the same precision of the prediction, as well as the same statistical distribution of the predictor coefficients.

4. Results

The first analysis concerns the distribution of the optimal coefficients for the linear predictor of the non-bonding forces. It is interesting to relate these predictor coefficients to those describing "perfect prediction" from an infinite time series in the past [4]. Any function $f(t)$ whose Fourier spectrum is bandwidth limited, such that $\tilde{f}(\omega) \equiv 0$ for $|\omega| > \omega_c$, can be exactly predicted from an infinite set equidistantly sampled values in the past if the sampling step fulfills the condition

$$\Delta t < \frac{\pi}{3\omega_c}. \tag{9}$$

In this case one can write

$$f(t) = \lim_{P\to\infty} \sum_{n=1}^{P} (-1)^{n+1} \binom{P}{n} f(t - n\Delta t). \tag{10}$$

This expression corresponds to a Taylor series where all the derivatives are replaced by corresponding differences constructed from the available time series for $f(t)$. For finite P Eq. (10) is a special case of (1), with the predictor coefficients

$$a_n = (-1)^{n+1} \binom{P}{n}, \quad n = 1, \ldots, P. \tag{11}$$

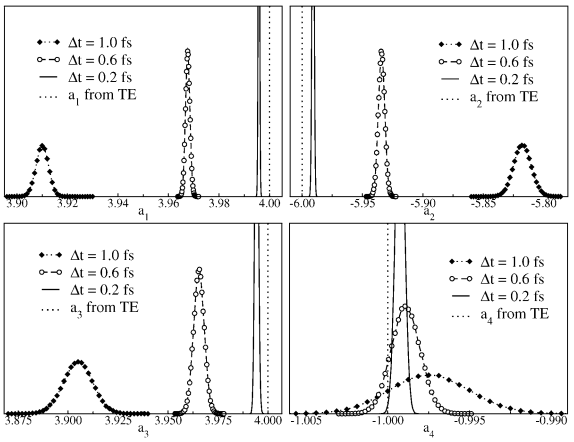

Fig. 3. Optimal predictor coefficients as a function of the time step, Δt. For $\Delta t \to 0$ all coefficients approach those of the Taylor extrapolation (TE). Here, the number of observations is $M = 15\,000$. To create the distributions shown above 100 000 **F** matrices have been randomly generated for each sampling time step. Here the order of prediction $P = 4$.

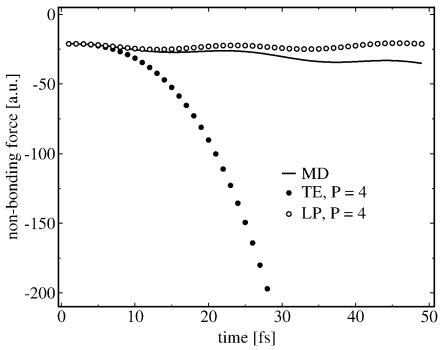

Fig. 4. Typical comparison of atomic non-bonding force trajectories obtained from linear prediction (LP), Taylor extrapolation (TE) and full MD for different integration steps ($\Delta t = 1$ fs).

Fig. 3 shows that the predictor coefficients obtained from (6) approach the Taylor coefficients (11) in the limit $\Delta t \to 0$.

Fig. 4 demonstrates that for typical simulation time steps the force trajectories obtained from linear prediction are in better agreement with the corresponding MD trajectories than those extrapolated by a Taylor series.

Figs. 5 and 6 show clearly that linear prediction produces atomic forces which are much closer to reality than Taylor extrapolation. The figures depict, respectively, the velocity distribution function and the

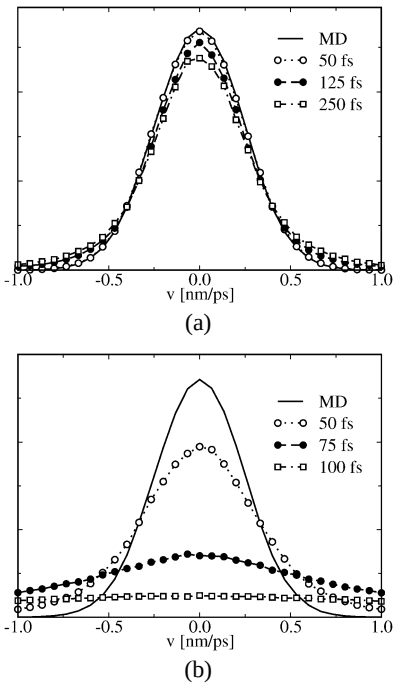

Fig. 5. Velocity distribution functions after a short simulation with, respectively, pure linear prediction (a) and pure Taylor extrapolation (b) of the non-bonding forces, compared to the reference distribution from full MD simulation.

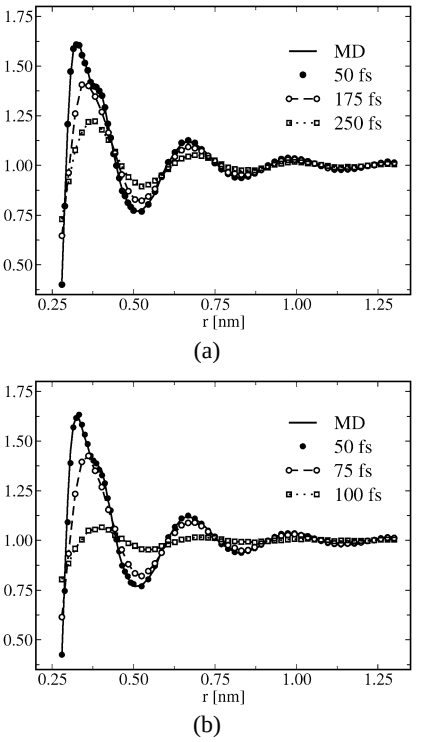

Fig. 6. Radial distribution functions after a short simulation with, respectively, pure linear prediction (a) and pure Taylor extrapolation (b) of the non-bonding forces, compared to the reference distribution from full MD simulation.

radial distribution function computed after short simulations with pure linear prediction ($P = 4$), pure Taylor extrapolation ($P = 4$), and the corresponding quantities from MD simulation. The velocity distribution function and the radial distribution function deviate much more slowly from the reference functions obtained by MD if linear prediction is used instead of Taylor extrapolation.

5. Conclusion

We have shown that linear prediction of intermolecular forces can be intertwined with the explicit calculation of the latter, in order to accelerate molecular dynamics simulations. Empirically, the predictor coefficients tend to those of a discrete Taylor expansion, if the time step tends to zero. We have shown that the deviation of the velocity distribution function and the radial distribution function from the respective reference quantities obtained from MD simulation is much slower for linear prediction than for Taylor expansion. Compared to the Taylor coefficients, the coefficients of a linear predictor are tailored for *finite* times steps and *small* orders P.

Acknowledgements

B.B. acknowledges financial support by a Marie Curie Reintegration Grant MERG-CT-2004-506291 within the EC Sixth Framework Programme.

References

[1] M. Allen, D. Tildesley, Computer Simulation of Liquids, Oxford University Press, Oxford, 1987.
[2] W. Streett, D. Tildesley, G. Saville, Multiple time step methods in molecular dynamics, Mol. Phys. 35 (1978) 639–648.
[3] B. Brutovsky, T. Mülders, G.R. Kneller, Accelerating molecular dynamics simulations by linear prediction of time series, J. Chem. Phys. 118 (2003) 6179–6187.
[4] A. Papoulis, Signal Analysis, McGraw-Hill, 1984.