Accelerating *Ab Initio* Molecular Dynamics Simulations by Linear Prediction Methods

Jonathan D. Herr, Ryan P. Steele*

Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, UT 84112

Henry Eyring Center for Theoretical Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, UT 84112

ARTICLE INFO

Article history:

Received XX July 2016

In final form XX July 2016

Available online XX July 2016

Keywords:

Ab initio Molecular Dynamics

Fock Extrapolation

Linear Prediction

ABSTRACT

Acceleration of *ab initio* molecular dynamics (AIMD) simulations can be reliably achieved by extrapolation of electronic data from previous timesteps. Existing techniques utilize polynomial least-squares regression to fit previous steps' Fock or density matrix elements. In this work, the recursive Burg "linear prediction" technique is shown to be a viable alternative to polynomial regression, and the extrapolation-predicted Fock matrix elements were three orders of magnitude closer to converged elements. Accelerations of 1.8-3.4x were observed in test systems, and in all cases, linear prediction outperformed polynomial extrapolation. Importantly, these accelerations were achieved without reducing the MD integration timestep.

1. Introduction

Ab initio molecular dynamics (AIMD) simulation techniques employ 'on-the-fly' generation of quantum chemistry molecular forces to propagate classical molecular dynamics trajectories on the Born-Oppenheimer potential energy surface. The ability to accurately capture strong polarization, charge transfer, and bond-breaking/-making events motivates the use of such techniques in chemical simulations of real-time dynamics, equilibrium sampling, and spectroscopic response. However, the many-orders-of-magnitude cost increase of AIMD, relative to commonly used molecular mechanics force fields, makes such simulations a daunting computational challenge for complex molecular systems.

Density-functional theory[1] (DFT) is the most commonly used quantum chemistry method for large AIMD simulations, although Møller-Plesset perturbation theory[2] (MP2)-based simulations have also been performed. For Kohn-Sham DFT or the underlying Hartree-Fock (HF) calculation for MP2, a non-linear, self-consistent field (SCF) optimization procedure is solved iteratively,[3] leading to an overall computational prefactor that is determined by the number of SCF cycles required for convergence to a chosen tolerance. Although advanced integral evaluation techniques[4,5] tackle the scaling of these simulations with respect to the size of the system, motivation remains to reduce the SCF-based prefactor,

particularly for molecular systems that are large, open-shell, and/or undergoing bond-rearrangement events.

The small timesteps required for AIMD trajectories (≤ 0.5 fs for hydrogen-containing systems) suggest that the bulk of the electronic structure has not appreciably changed between neighboring timesteps. Restarting the SCF procedure anew from a typically poor initial guess, therefore, would appear to be computationally wasteful. This notion has motivated two diverging routes to accelerating MD simulations.

One option is the *propagation* of electronic information to the next timestep. Extended-Lagrangian MD (ELMD) methods[6-8] accelerate—or altogether avoid—the SCF procedure by propagating Fock or density matrix information. Fidelity with the true Born-Oppenheimer surface is approximated through the extended-Lagrangian formalism. One often-cited advantage of this approach is the rigorous enforcement of total energy conservation. Tradeoffs include the need for typically smaller MD timesteps (thereby reducing the effectiveness of the computational acceleration), choice of a fictitious mass, and the potential for electronic 'lag' artifacts[8-10] that can appear in observables, such as vibrational spectra. More recent implementations[11-16] successfully bridge the ELMD-BOMD divide, although stable propagation of the electronic variables for reactive systems remains an area of active research. One of the clear advantages to time-

reversibly propagating electronic information has been the demonstration by Niklasson that the SCF procedure does not need to be converged tightly—sometimes requiring only a single SCF step [13,15]—in order to produce energy-conserving trajectories.

The second option involves *extrapolation* of electronic information from previous timesteps. Motivated by the observation that Fock matrix elements, for example, are simple oscillatory functions of time, these approaches attempt to fit previous timesteps' information to a simple form that can be used as an improved initial guess for the SCF procedure.[17-20] This approach allows for longer timesteps while still reducing the number of requisite SCF cycles. In its simplest version, this approach uses the converged Fock matrix from the previous timestep as this initial guess, which is consistent with the intuitive notion that the electronic structure has not changed appreciably. However, as has been noted in previous analyses,[18] such a choice quickly fails when loose SCF convergence settings are employed. The use of the previous step's Fock matrix introduces a hysteresis effect that leads to rather severe non-conservation of the energy. Over only short trajectories, the energy drift swamps the inherent energy fluctuations of the chosen integration scheme. While higher-order polynomials somewhat dampen this energy-conservation concern, its effects should eventually be borne out in long MD trajectories since the history-dependent guess renders the propagator formally non-symplectic. The practical cure to this pathology, of course, is to more tightly converge the SCF procedure. As will be shown below, this approach rectifies energy-drift concerns for very long trajectories (by AIMD standards), but it also necessarily leads to additional computational cost overhead.

The single SCF cycle performed in the 'propagation' approach could potentially lead to unstable electronic states for slow-to-converge SCF cases, such as open-shell ions or systems near bond-arrangement events, and these cases require additional testing in the literature. Because of our interest in these difficult systems,[21] the present analysis focuses solely on the 'extrapolation' approach, with the intent of improving existing acceleration techniques within tight SCF convergence settings. An investigation into whether the behavior observed in the present analysis can be combined with rigorously time-reversible propagators is reserved for future studies.

Existing Fock extrapolation techniques employ a polynomial least-squares regression (PLSR) of N previous timesteps' information, to a polynomial of a chosen order O . A recent analysis[22] has shown that an optimal N may exist, at least for a given timestep and chemical system, although no assessment of O -dependence was performed. Nonetheless, past studies have shown that with only a handful of previous steps' data, roughly 2-fold reductions in the number of Fock builds can be achieved. Extension to MP2 simulations, in both the SCF and gradient response terms, has also been demonstrated.[23]

While this standard PLSR approach is appropriate for accurately fitting the data within the chosen time window, it is decidedly *not* optimal for extrapolating information outside of this window. Given that the Fock data throughout an MD trajectory is locally oscillatory (with some slight noise, stemming from incomplete convergence) and globally stochastic, techniques that are well-suited to signal prediction were conjectured to be a better choice. In particular, so-called *linear prediction* techniques—not to be confused with linear PLSR extrapolation ($O=1$)—are designed to accurately represent exactly this type of behavior. This study presents an implementation of the Burg Linear Prediction (BLP) algorithm for Fock extrapolation, as well as rigorous tests of its efficacy and efficiency for accelerating AIMD simulations.

2. Methods

The basic components of the BLP algorithm are presented in this section, including a discussion of the distinguishing features of the BLP and PLSR approaches. This presentation is followed by a description of the techniques used to assess the methods throughout the remainder of the study.

2.1 Burg Linear Prediction

The BLP algorithm,[24,25] in an extrapolated AIMD context, fits N previous timesteps' atomic-orbital (AO) basis Fock matrix elements to a *recursive series* with L coefficients forward [f] in time:

$$\tilde{F}_{\mu\nu}^{[f]}(t) = \sum_{i=1}^L c_i F_{\mu\nu}(t-i) \quad (1)$$

The Burg algorithm augments this simpler Levinson-Durbin recursion definition by also considering a fit to backward [b] predictions:

$$\tilde{F}_{\mu\nu}^{[b]}(t) = \sum_{i=1}^L c_i F_{\mu\nu}(t+i) \quad (2)$$

In this notation, the predicted element is denoted as $\tilde{F}$, and t acts as a timestep index, rather than a physical simulation time. In all analyses performed in this work, the timesteps were evenly spaced, although the LPB prescription is not restricted to this form. The coefficients are obtained using linear least-squares minimization of the *sum* of the forward and backward mean-square deviations, χ^2 , defined over the fitting length N as

$$\chi_{\mu\nu}^2 [f] = \sum_{t=L}^N [F_{\mu\nu}(t) - \tilde{F}_{\mu\nu}^{[f]}(t)]^2 = \sum_{t=L}^N \left[F_{\mu\nu}(t) - \left(\sum_{i=1}^L c_i F_{\mu\nu}(t-i) \right) \right]^2 \quad (3)$$

$$\chi_{\mu\nu}^2 [b] = \sum_{t=L}^N [F_{\mu\nu}(t) - \tilde{F}_{\mu\nu}^{[b]}(t)]^2 = \sum_{t=L}^N \left[F_{\mu\nu}(t) - \left(\sum_{i=1}^L c_i F_{\mu\nu}(t+i) \right) \right]^2 \quad (4)$$

This optimization of $\chi^2 = \chi^2 [f] + \chi^2 [b]$, rather than $\chi^2 [f]$ alone, distinguishes BLP from simpler recursion approaches and introduces considerable stability to the method. Further algorithmic details can be found in standard numerical texts, including the one used in the present implementation.[24]

The BLP algorithm, therefore, generates a weighted linear combination of existing Fock matrix elements. In this sense, it is superficially similar to a PLSR fit,

$$\tilde{F}_{\mu\nu}^{[f]}(t) = \sum_{i=0}^O c_i \cdot t^i \quad (5)$$

However, as will be shown below, the BLP algorithm was found to be much more powerful for extrapolation purposes. For sums of Fourier signals, for example, preliminary tests of this approach yielded nearly quantitative reproduction of the signal for thousands of extrapolated signal cycles, rather than the 1-2 that was afforded with PLSR. Of course, in an AIMD context, the fit will be recomputed at each timestep, meaning that the algorithm only needs to be accurate to the next timestep. Even in this regime, though, the BLP approach led to several-orders-of-magnitude reductions in error for values of this subsequent step.

The use of PLSR requires a choice of N and O , compactly notated in the remainder of this work as an (N, O) extrapolation. Commonly used choices in previous studies[17,18] included (4,2) to (20,10) extrapolations. Since this approach is intended as an extrapolation,

rather than a fit, increases in N or O do not necessarily yield better SCF guesses, and the optimal choice is often simulation-specific.[22] For BLP, analogous choices must be made, and the notation will similarly be defined as an (N, L) extrapolation. In practice, fits to a larger number of data points N (10-100) were required in BLP than in PLSR, but the subsequent extrapolation behavior was also much less sensitive to these parameters, as will be shown below. For atom-centered basis sets, no appreciable overhead is incurred by either approach, compared to the overwhelming cost of computing *ab initio* energies and forces.

2.2 Numerical Analyses

The guiding question to be addressed in this analysis was whether the BLP algorithm led to any appreciable reduction in the requisite number of SCF cycles for AIMD simulations. Microcanonical AIMD simulations were mainly performed on two hydrogen-containing test systems, in order to address this issue. Small systems were selected in order to enable testing of the long-time energy conservation properties of these trajectories. A single water molecule, H_2O , represented the regime in which the SCF was already reasonably simple to converge. The lack of a large density of low-frequency vibrational modes actually made this case challenging for extrapolation techniques, however. The water dimer cation, $(H_2O)_2^+$ (more accurately described as $H_3O^+ \cdots OH^+$), represented the difficult-to-converge regime. Past experience with this open-shell ion highlighted the challenge in converging its SCF—often requiring 20-50 cycles for even this small complex—and its use here serves as a small-molecule model of challenges encountered in larger systems. Subsequently, the larger diglycine molecule, $C_4N_2O_3H_8$, was examined, in order to ensure that results were not specific to these small water complexes. For testing purposes, most of the simulations employed the HF method with the 6-31G(d,p) basis set,[26,27] although the conclusions are likely invariant to this choice. The dipeptide was tested with B3LYP-D[28-30]/cc-pVTZ[31]. All computations were performed in a development version of the Q-Chem 4.3 quantum chemistry software package.[32]

Several auxiliary factors influence the cost and stability of extrapolated AIMD trajectories, including the extrapolation parameters, the MD integration timestep, and the degree of SCF convergence. Tests that varied parameters along all three of these axes were performed on the two small test complexes:

(1) The influence of the number of BLP points (N) was examined by holding the number of coefficients (L) constant at 10 while increasing the number of points from 11 to 100, by orders of 10. The L dependence was subsequently examined by holding the number of points constant at 100 and increasing the number of coefficients from 10 to 99, also by orders of ten.

(2) The behavior of the extrapolation schemes was also examined as a function of MD integration timestep, Δt . Two timesteps—10 a.u. (0.242 fs) and 20 a.u. (0.484 fs)—were examined in this analysis, which will hereafter be referred to as Δt_a and Δt_b , respectively. The velocity Verlet integrator was used in all tests.

(3) Dependence on the SCF convergence criterion was performed with a (100,30) BLP extrapolation using SCF convergence settings of $\{10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}\}$ a.u. In these analyses, the convergence criterion was defined as either the RMS value of the DIIS[33,34] error vector or the maximum value of the energy-weighted GDM[35] orbital gradient, depending on the convergence algorithm employed.

The main metric of interest was the trajectory-averaged number of SCF cycles. In all cases, statistics were only generated after the

extrapolation period began. Also examined were the RMS total relative energy fluctuations, defined in this work as

$$\Delta = \log_{10} \left[\frac{\langle (E(t_i) - E(t_0))^2 \rangle^{\frac{1}{2}}}{E(t_0)} \right] \quad (6)$$

where the average was computed over the entire trajectory. In order to assess any energy non-conservation (drift), the energy fluctuations were also fit to a line, and the corrected fluctuations, Δ^{corr} , with this linear fit subtracted were also computed analogously to eqn (6).

Comparison to non-extrapolated trajectories, with the same settings, was performed throughout. Initial conditions included equilibrium structures and momenta selected from a Maxwell-Boltzmann distribution at 300K. Once selected, these initial conditions were fixed for all NVE trajectories in a given analysis, in order to provide an even-handed comparison of methods. The SCF-step analysis was performed on 5-ps trajectories (roughly 10,000-20,000 MD steps), and trajectory stability analyses were performed on 45-ps trajectories. In all cases, an integral drop tolerance[36,37] of 10^{-12} a.u. was employed.

3. Results

3.1 A Demonstrative Example

Before presenting statistical averages across the suite of tests, an explicit example was chosen to highlight the utility of the BLP approach. Using H_2O with Δt_b (20 a.u.), a selected single Fock matrix element was examined. This element was chosen as one that oscillated significantly on the $O-H$ stretch timescale, although extrapolation behavior for all elements was found to be similar. Figure 1(a) shows the computed value of this element across a 5,000-step trajectory. A 100-step section of this trajectory is shown in Fig 1(b), where the value of the extrapolated-guess Fock elements for BLP (100,30) and PLSR (16,8) methods are also shown. Visually, the difference between the two methods can be seen in panel (b). The BLP guess tightly followed the converged Fock result, whereas the PLSR showed more scatter about the converged value. To better quantify this difference, the deviation between the extrapolated guess and the converged Fock matrix element is shown in Fig 1(c), across the trajectory (after extrapolation began). The BLP approach yielded guess values that were more than three orders of magnitude closer to the converged value, which suggested that this approach should lead to improved SCF performance.

3.2 Linear Prediction Parameters and Timestep Analysis

The results of the various (N, L) parameter sets for both small test systems are listed in Tables 1 and 2. The overall results in Table 1 indicate that BLP extrapolation provided appreciable reductions in the average number of SCF cycles. Including more previous data points in the extrapolation, when using $L=10$ coefficients, generally increased the effectiveness of the extrapolation, although the results became essentially saturated near 50 points. At this setting, average SCF accelerations ranged from $1.8\times$ ($H_2O, \Delta t=0.484$ fs) to $3.2\times$ ($(H_2O)_2^+, \Delta t=0.242$ fs). The extrapolation was, not surprisingly, less effective at the longer timestep, but cost reductions were still appreciable at this most aggressive integration setting. The cost reductions were considerably more impressive in the tough-to-converge water dimer cation case.

Fig. 1 Explicit Fock matrix element behavior for H_2O with $\Delta t=20$ a.u. (a) Single Fock matrix element value across the trajectory. (b) 100-step close-up of converged Fock element, as well as PLSR (red) and BLP (blue) extrapolated guess Fock element values. (c) Absolute deviation between extrapolated guess values and converged values. Colors are consistent with (b).

The results in Table 2 show that the coefficient dependence of BLP extrapolation, when the number of points is held fixed ($N=100$), is rather small. Although each test exhibited an optimum value of this parameter, the fluctuations in average SCF cycles as a function of this parameter were negligible compared to the magnitude of the

Table 1

SCF cycles, averaged over 5-ps AIMD trajectories with HF/6-31G(d,p). Shown are results with and without BLP extrapolation. Results for two integration timesteps are listed: $\Delta t_a=0.242$ fs, $\Delta t_b=0.484$ fs. SCF convergence was set to 10^{-8} a.u.

Extrapolation Parameters		Average SCF Cycles			
		H_2O		$(H_2O)_2^+$	
Points	Coefficients	Δt_a	Δt_b	Δt_a	Δt_b
11	10	8.87	10.10	16.87	18.78
20	10	7.08	8.97	13.73	16.34
30	10	5.79	8.68	12.09	15.09
40	10	5.33	8.50	11.28	14.69
50	10	5.20	8.43	10.74	14.35
60	10	5.16	8.42	10.54	14.41
70	10	5.11	8.42	10.69	14.44
80	10	5.12	8.42	10.25	14.44
90	10	5.12	8.41	10.29	14.37
100	10	5.12	8.42	10.20	14.36
(No Extrapolation)		14.77	14.76	34.81	35.00

Table 2

SCF cycles, averaged over 5-ps AIMD trajectories with HF/6-31G(d,p). Shown are results with and without BLP extrapolation. Results for two integration timesteps are listed: $\Delta t_a=0.242$ fs, $\Delta t_b=0.484$ fs. SCF convergence was set to 10^{-8} a.u.

Extrapolation Parameters		Average SCF Cycles			
		H_2O		$(H_2O)_2^+$	
Points	Coefficients	Δt_a	Δt_b	Δt_a	Δt_b
100	10	5.12	8.42	10.20	14.36
100	20	4.80	7.99	9.79	14.06
100	30	4.76	7.94	9.57	13.94
100	40	4.75	7.89	9.69	13.78
100	50	4.78	7.84	9.81	14.12
100	60	4.81	7.85	9.92	14.36
100	70	4.84	7.87	9.84	14.42
100	80	4.85	7.91	9.87	14.17
100	90	4.87	7.93	10.05	14.41
100	99	4.88	7.95	10.17	14.20
(No Extrapolation)		14.77	14.76	34.81	35.00

overall reduction. This result, combined with the saturation shown in Table 1, was particularly encouraging since the acceleration appeared to be much less sensitive to the choice of extrapolation parameters than PLSR methods. For example, a (4,2) PLSR extrapolation for $(H_2O)_2^+$ at $\Delta t=10$ a.u. (0.242 fs) required an average of 18.18 SCF cycles. An alternative (16,8) extrapolation improved this result to only 11.39 cycles. However, further increasing the order to a (20,10) extrapolation yielded 16.11 cycles. Therefore, once beyond low-order extrapolations, the BLP approach appeared to be sufficiently resilient to choice of parameters, and a roughly (100,30) extrapolation is recommended, based on this data.

3.3 SCF Convergence Analysis

The number of requisite SCF cycles is naturally tied to the degree of convergence in AIMD simulations. The performance of Fock extrapolation techniques was analyzed at several convergence settings. Using a fixed (100,30) LPB extrapolation, the number of SCF cycles was shown to increase with convergence setting, as expected; results are shown in Table 3. In all cases, however, the LPB approach provided significant acceleration over both extrapolation-free trajectories, as well as PLSR. For the latter method, a crude optimization of the extrapolation order O (with N fixed at $2 \cdot O$) was performed, and the results shown in Table 3

Table 3

SCF cycles, averaged over 5-ps AIMD trajectories with HF/6-31G(d,p). Shown are results with and without PLSR/BLP extrapolation for several SCF convergence criteria (Conv=10⁻ⁿ). Results for two integration timesteps are listed: $\Delta t_a=0.242$ fs, $\Delta t_b=0.484$ fs. The BLP results employed a (100,30) extrapolation throughout; the PLSR results used the optimal extrapolation, among (20,10), (16,8), (12,6), (8,4), and (4,2) extrapolation options, as a 'best case' comparison for each system.

System	Method	Average SCF Cycles			
		Conv=4	6	8	10
H_2O Δt_a	BLP	1.78	3.01	4.76	6.61
	PLSR	3.18	4.46	7.52	10.78
	None	9.00	11.80	14.77	17.70
H_2O Δt_b	BLP	2.46	4.93	7.84	11.13
	PLSR	4.53	7.56	10.82	14.23
	None	9.00	11.80	14.76	17.70
$(H_2O)_2^+$ Δt_a	BLP	2.89	6.26	9.57	13.57
	PLSR	9.25	9.70	11.39	17.45
	None	24.75	30.58	34.81	42.17
$(H_2O)_2^+$ Δt_b	BLP	4.79	9.02	13.78	18.01
	PLSR	10.01	12.21	18.10	22.91
	None	24.51	30.39	35.00	42.67

represent the best-performing PLSR setting. Complete data tables can be found in the Supporting Information. The difference between the two extrapolation techniques was more pronounced at shorter timesteps, but BLP still provided non-negligible savings over PLSR at the longer of the two timesteps.

3.4 Energy Conservation Analysis

As mentioned in Sec 1, any extrapolation of previous timestep data formally breaks time-reversibility in AIMD simulations. This behavior is clearly evident in the plot of energy fluctuations shown in Fig 2(a), for which the SCF convergence setting (DIIS) was set to 10⁻⁵ a.u. ($\frac{E_h}{a_0}$) and $\Delta t_b=20$ a.u. Without extrapolation, the BOMD trajectory exhibited long-term stability. Energy fluctuations, on the

order of 0.1 milliHartree, were exhibited, but no observable drift in the energy was found. Conversely, both PLSR and BLP exhibited drifts. By the end of the 45-ps trajectory, the drift in the PLSR trajectory had surpassed the inherent fluctuations stemming from the velocity Verlet integrator. The drift in the BLP results was smaller but clearly still present.

Encouragingly, measurable drifts in all methods were removed with the SCF convergence tightened to 10⁻⁸. In Fig 2(b), all three methods overlapped, exhibiting equal energy fluctuation magnitudes ($\Delta=-4.058$, see eqn (6)). The average slope of all three methods was <10⁻¹¹ Hartree/step, which was within the bounds of endpoint effects in the linear fit for this trajectory length. Therefore, tight SCF convergence can practically remove any energy-conservation concerns, even for relatively long trajectories. Of course, this tight tolerance also increases the number of SCF cycles. As was demonstrated in Table 3, BLP speedups are still appreciable in this tight convergence window. In particular, the required number of BLP SCF cycles was always lower at a convergence setting of 10⁻⁸ than for non-extrapolated trajectories at a setting of 10⁻⁴.

3.5 Canonical Sampling

One additional usage of AIMD is for the purpose of sampling thermal equilibrium ensembles. To our knowledge, no explicit analysis of the efficacy of extrapolation methods in this context was available in the literature. Therefore, the above tests were repeated in the NVT ensemble, using a white-noise Langevin thermostat (time constant = 100 fs). The average number of SCF cycles from these trajectories are compared in Table 4. For H_2O (Δt_b), the BLP extrapolation is less effective in this regime; the stochastic 'kicks' imparted by the thermostat increase the number of SCF cycles from 7.8 to 10.4 (14.8 without extrapolation). Similar trends were observed for $(H_2O)_2^+$. These tests represent a worst-case scenario because the velocity alterations imparted by this very tightly coupled thermostat most strongly affect small systems. In larger systems, where energy can more easily redistribute naturally and where looser thermostat coupling can be employed, SCF behavior closer to the microcanonical result is anticipated.

Fig. 2 Energy drift analysis for H_2O with $\Delta t=20$ a.u. Shown are results for extrapolation-free trajectories (blue), BLP extrapolation (green), and PLSR (red). **(a)** Energy fluctuations with an SCF (DIIS) convergence of 10⁻⁵ a.u. In this panel, the extrapolated values drift, with roughly constant fluctuation magnitude relative to this drift (data behind the no-extrapolation results). **(b)** Energy fluctuations with an SCF convergence of 10⁻⁸ a.u. All data were nearly identical and overlap in this panel.

Table 4

SCF cycles, averaged over 5-ps AIMD trajectories with HF/6-31G(d,p), $\Delta t=20$ a.u., and SCF convergence of 10^{-8} a.u. Shown are results with and without PLSR/BLP extrapolation for canonical (NVT) and microcanonical (NVE) simulations. The former used a white-noise, Langevin thermostat. The BLP results employed a (100,30) extrapolation throughout; the PLSR results used the optimal extrapolation, among (20,10), (16,8), (12,6), (8,4), and (4,2) extrapolation options, as a 'best case' comparison for each system.

System	Method	Average SCF Cycles	
		NVE	NVT
H_2O	BLP	7.84	10.36
	PLSR	10.82	10.86
	None	14.76	14.83
$(H_2O)_2^+$	BLP	13.78	21.07
	PLSR	18.10	21.34
	None	35.00	30.39

3.6 Larger Test System

The larger diglycine test system was also examined, using a B3LYP-D/cc-pVTZ potential. With an MD integration timestep of $\Delta t=15$ a.u., a non-extrapolated BOMD trajectory required an average of 15.0 SCF cycles for convergence to 10^{-8} a.u. The PLSR extrapolation reduced this cost to an average of 6.9 cycles, whereas the BLP further reduced the cost to 5.0 cycles. Therefore, both approaches provided significant acceleration in this larger system, and BLP once again outperformed the PLSR approach. The fact that BLP is also appropriate for DFT-based potentials and with large basis sets is particularly encouraging, given that promising new density functionals (particularly those with non-local components) often require large basis sets.[38]

4. Conclusions

This study has presented an implementation of the Burg Linear Prediction algorithm for the purpose of enhancing Fock-extrapolated AIMD simulations. The BLP algorithm fits a recursive series to previous timesteps' Fock matrix data and is tailor-made for

extrapolating (slightly noisy) oscillatory signals. This new approach was shown to be superior to existing polynomial least-squares regression techniques, in terms of reducing the number of requisite SCF cycles during AIMD trajectories. Like any extrapolation method, it still formally breaks time-reversal symmetry. However, this new method provides a route to accelerating tightly converged SCF-based dynamics, and tests of long trajectories showed no appreciable energy drift. Furthermore, the results were notably less sensitive to the choice of extrapolation parameters, which should make this new approach a reliable alternative across disparate chemical systems.

The BLP extrapolation typically uses more of the previous timesteps' data, which means that the extrapolation 'warmup' period is longer. For long trajectories, this difference is negligible. For shorter trajectories, the relative insensitivity of the method to extrapolation parameters suggests that the extrapolation could, in principle, actually begin after only a few steps (similar to polynomial extrapolation), and the traveling extrapolation window could be augmented until the full window is filled. This approach would also allow for on-the-fly saturation tests, in order to automatically determine this optimal window size.

Implementation of the BLP method in quantum chemistry programs is actually very similar to machinery used for polynomial regression, and our present implementation recycled much of the existing code. Based on the above results and this simple implementation, we expect that linear prediction-based extrapolation could become a direct replacement in existing AIMD codes.

Acknowledgements

This material is based on work supported by the United States National Science Foundation CAREER under CHE-1452596. The support and resources from the Center for High-Performance Computing at the University of Utah are gratefully acknowledged. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1053575.

References

- [1] D.S. Sholl, J.A. Steckel, *Density Functional Theory: A Practical Introduction*, Wiley-Interscience, 2009.
- [2] C. Møller, M.S. Plesset, *Phys. Rev.* 46 (1934) 618.
- [3] A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill, Inc., New York, 1989.
- [4] C.A. White, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, *Chem. Phys. Lett.* 230 (1994) 8.
- [5] C. Ochsenfeld, J. Kussmann, D.S. Lambrecht, *Rev. Comput. Chem.*, John Wiley & Sons, Inc., 2007, p. 1.
- [6] R. Car, M. Parrinello, *Phys. Rev. Lett.* 55 (1985) 2471.
- [7] H.B. Schlegel, J.M. Millam, S.S. Iyengar, G.A. Voth, A.D. Daniels, G.E. Scuseria, M.J. Frisch, *J. Chem. Phys.* 114 (2001) 9758.
- [8] J.M. Herbert, M. Head-Gordon, *J. Chem. Phys.* 121 (2004) 11542.
- [9] P. Tangney, S. Scandolo, *J. Chem. Phys.* 116 (2002) 14.
- [10] I.F.W. Kuo, C.J. Mundy, M.J. McGrath, J.I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M.L. Klein, F. Mohamed, M. Krack, M. Parrinello, *J. Phys. Chem. B* 108 (2004) 12990.
- [11] A.M.N. Niklasson, C.J. Tymczak, M. Challacombe, *Phys. Rev. Lett.* 97 (2006) 123001.
- [12] M.J. Cawkwell, A.M.N. Niklasson, *J. Chem. Phys.* 137 (2012) 134105.
- [13] P. Souvatzis, A.M.N. Niklasson, *J. Chem. Phys.* 139 (2013) 214102.
- [14] A.M.N. Niklasson, M.J. Cawkwell, *J. Chem. Phys.* 141 (2014) 164123.
- [15] P. Souvatzis, A.M.N. Niklasson, *J. Chem. Phys.* 140 (2014) 044117.
- [16] T.D. Kühne, M. Krack, F.R. Mohamed, M. Parrinello, *Phys. Rev. Lett.* 98 (2007) 066401.
- [17] P. Pulay, G. Fogarasi, *Chem. Phys. Lett.* 386 (2004) 272.
- [18] J.M. Herbert, M. Head-Gordon, *Phys. Chem. Chem. Phys.* 7 (2005) 3269.
- [19] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Comput. Phys. Commun.* 167 (2005) 103.
- [20] T. Atsumi, H. Nakai, *J. Chem. Phys.* 128 (2008) 094101.
- [21] J.D. Herr, J. Talbot, R.P. Steele, *J. Phys. Chem. A* 119 (2015) 752.
- [22] J. Fang, X. Gao, H. Song, H. Wang, *J. Chem. Phys.* 144 (2016) 244103.
- [23] R.P. Steele, J.C. Tully, *Chem. Phys. Lett.* 500 (2010) 167.
- [24] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 2002.
- [25] J.P. Burg, in: D.G. Childers (Ed.), *Modern Spectrum Analysis*, IEEE Press.
- [26] W.J. Hehre, R.F. Stewart, J.A. Pople, *J. Chem. Phys.* 51 (1969) 2657.
- [27] W.J. Hehre, R. Ditchfield, J.A. Pople, *J. Chem. Phys.* 56 (1972) 2257.
- [28] A.D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [29] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, *J. Phys. Chem.* 98 (1994) 11623.
- [30] S. Grimme, *J. Comput. Chem.* 27 (2006) 1787.
- [31] T.H. Dunning, *J. Chem. Phys.* 90 (1989) 1007.
- [32] Y. Shao, Z. Gan, E. Epifanovsky, A.T.B. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P.R. Horn, L.D. Jacobson, I. Kaliman, R.Z. Khaliullin, T. Kuš, A. Landau, J. Liu, E.I. Proynov, Y.M. Rhee, R.M. Richard, M.A. Rohrdanz, R.P. Steele, E.J. Sundstrom, H.L. Woodcock, P.M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G.J.O. Beran, Y.A. Bernard, E. Berquist, K. Brandhorst, K.B. Bravaya, S.T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S.H. Chien, K.D. Closser, D.L. Crittenden, M. Diedenhofen, R.A. DiStasio, H. Do, A.D. Dutoi, R.G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M.W.D. Hanson-Heine, P.H.P. Harbach, A.W. Hauser, E.G. Hohenstein, Z.C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R.A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C.M. Krauter, K.U. Lao, A.D. Laurent, K.V. Lawler, S.V. Levchenko, C.Y. Lin, F. Liu, E. Livshits, R.C. Lochan, A. Luenser, P. Manohar, S.F. Manzer, S.-P. Mao, N. Mardirossian, A.V. Marenich, S.A. Maurer, N.J. Mayhall, E. Neuscamman, C.M. Oana, R. Olivares-Amaya, D.P. O'Neill, J.A. Parkhill, T.M. Perrine, R. Peverati, A. Prociuk, D.R. Rehn, E. Rosta, N.J. Russ, S.M. Sharada, S. Sharma, D.W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A.J.W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M.A. Watson, J. Wenzel, A. White, C.F. Williams, J. Yang, S. Yeganeh, S.R. Yost, Z.-Q. You, I.Y. Zhang, X. Zhang, Y. Zhao, B.R. Brooks, G.K.L. Chan, D.M. Chipman, C.J. Cramer, W.A. Goddard, M.S. Gordon, W.J. Hehre, A. Klamt, H.F. Schaefer, M.W. Schmidt, C.D. Sherrill, D.G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A.T. Bell, N.A. Besley, J.-D. Chai, A. Dreuw, B.D. Dunietz, T.R. Furlani, S.R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D.S. Lambrecht, W. Liang, C. Ochsenfeld, V.A. Rassolov, L.V. Slipchenko, J.E. Subotnik, T. Van Voorhis, J.M. Herbert, A.I. Krylov, P.M.W. Gill, M. Head-Gordon, *Mol. Phys.* 113 (2015) 184.
- [33] P. Pulay, *Chem. Phys. Lett.* 73 (1980) 393.
- [34] P. Pulay, *J. Comput. Chem.* 3 (1982) 556.
- [35] T. Van Voorhis, M. Head-Gordon, *Mol. Phys.* 100 (2002) 1713.
- [36] J. Almlöf, K. Faegri, K. Korsell, *J. Comput. Chem.* 3 (1982) 385.
- [37] M. Häser, R. Ahlrichs, *J. Comput. Chem.* 10 (1989) 104.
- [38] N. Mardirossian, M. Head-Gordon, *J. Chem. Theory Comput.* 9 (2013) 4453.