Accelerating Fermionic Molecular Dynamics

A D Kennedy
School of Physics
University of Edinburgh
Abstract

We consider how to accelerate Fermionic Molecular Dynamics algorithms by introducing $n$ pseudofermion fields coupled with the $n^{\text{th}}$ root of the Fermionic kernel. The $n^{\text{th}}$ roots may be computed efficiently using Чебышев rational approximations, and to this end we review the theory of optimal polynomial and rational Чебышев approximations.
Non-linearity of CG solver

- Suppose we want to solve $A^2x=b$ for Hermitian $A$ by CG
  - It is better to solve $Ax=y$, $Ay=b$ successively
  - Condition number $\kappa(A^2) = \kappa(A)^2$
  - Cost is thus $2\kappa(A) < \kappa(A^2)$ in general
- Suppose we want to solve $Ax=b$
  - Why don’t we solve $A^{1/2}x=y$, $A^{1/2}y=b$ successively?
- The square root of $A$ is uniquely defined if $A>0$
  - This is the case for fermion kernels
- All this generalises trivially to $n^{th}$ roots
  - No tuning needed to split condition number evenly
- How do we apply the square root of a matrix?
Rational matrix approximation

- Functions on matrices
  - Defined for a Hermitian matrix by diagonalisation
    \[ H = UDU^{-1} \]
    \[ f(H) = f(UDU^{-1}) = U f(D) U^{-1} \]
- Rational functions do not require diagonalisation
  - \[ \alpha H^m + \beta H^n = U(\alpha D^m + \beta D^n) U^{-1} \]
  - \[ H^{-1} = UD^{-1}U^{-1} \]
- Rational functions have nice properties
  - Cheap (relatively)
  - Accurate
Polynomial approximation

What is the best polynomial approximation \( p(x) \) to a continuous function \( f(x) \) for \( x \) in \([0, 1]\) ?

Best with respect to the appropriate norm

\[
\|p - f\|_n = \left( \int_0^1 dx |p(x) - f(x)|^n \right)^{1/n}
\]

where \( n \geq 1 \)
Weierstraß’ theorem

- Taking $n \to \infty$ this is the “minimax” norm

$$\| p - f \|_\infty = \min_p \max_{0 \leq x \leq 1} |p(x) - f(x)|$$

- Weierstraß: Any continuous function can be arbitrarily well approximated by a polynomial
The explicit solution is provided by Бернштейне polynomials

\[ p_n(x) \equiv \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^n (1-x)^{n-k} \]
Чебышев’s theorem

- Чебышев: There is always a unique polynomial of any degree $d$ which minimises
  \[ \| p - f \|_\infty = \max_{0 \leq x \leq 1} |p(x) - f(x)| \]

- The error $|p(x) - f(x)|$ reaches its maximum at at least $d+2$ points on the unit interval
Чебышев’s theorem: Necessity

- Suppose \( p-f \) has less than \( d+2 \) extrema of equal magnitude
- Then at most \( d+1 \) maxima exceed some magnitude
- This defines a “gap”
- We can construct a polynomial \( q \) of degree \( d \) which has the opposite sign to \( p-f \) at each of these maxima (Lagrange interpolation)

Lagrange was born in Turin!
Чебышев’s theorem: Sufficiency

- Suppose there is a polynomial $\|p' - f\| \leq \|p - f\|$
- Then $|p'(x_i) - f(x_i)| \leq |p(x_i) - f(x_i)|$
- Therefore $p' - p$ must have $d+1$ zeros on the unit interval
- Thus $p' - p = 0$ as it is a polynomial of degree $d$

125th Anniversary of Чебышев’s birth
Чебышев polynomials

- Convergence is often exponential in $d$
  - The best approximation of degree $d-1$ over $[-1,1]$ to $x^d$ is $p_{d-1}(x) \equiv x^d - \left(\frac{1}{2}\right)^{d-1} T_d(x)$
    - Where the Чебышев polynomials are $T_d(x) = \cos\left(d \cos^{-1}(x)\right)$
    - The notation is an old transliteration of Чебышев!
  - In general, truncated Чебышев $L^2$ approximation is not $L^\infty$ optimal even for polynomials
- The error is $\left\|x^d - p_d(x)\right\|_\infty = \left(\frac{1}{2}\right)^{d-1} \left\|T_d(x)\right\|_\infty = 2e^{-d \ln 2}$
Чебышев rational functions

- Чебышев’s theorem is easily extended to rational approximations
  - Rational functions with nearly equal degree numerator and denominator are usually best
  - Convergence is still often exponential
  - Rational functions usually give a much better approximation

- A simple (but somewhat slow) numerical algorithm for finding the optimal Чебышев rational approximation was given by Ремез
Чебышев rationals: Example

- A realistic example of a rational approximation is
  \[
  \frac{1}{\sqrt{x}} \approx 0.3904603901 \frac{(x + 2.3475661045)(x + 0.1048344600)(x + 0.0073063814)}{(x + 0.4105999719)(x + 0.0286165446)(x + 0.0012779193)}
  \]
- This is accurate to within almost 0.1% over the range [0.003,1]
- Using a partial fraction expansion of such rational functions allows us to use a multishift linear equation solver, thus reducing the cost significantly.
- The partial fraction expansion of the rational function above is
  \[
  \frac{1}{\sqrt{x}} \approx 0.3904603901 + \frac{0.0511093775}{x + 0.0012779193} + \frac{0.1408286237}{x + 0.0286165446} + \frac{0.5964845033}{x + 0.4105999719}
  \]
- This appears to be numerically stable.
Чебышев rationals: Example

- $\text{sgn}(x) = x \, R(x^2) + O(\Delta)$
- $\epsilon < |x| < 1$
- $\epsilon = 0.01$
- $R$ of degree $(10,10)$
- $\Delta = 0.00437985$
- Coefficients are known in closed form in terms of elliptic functions (Золотарев)
Золотарев’s formula: 1

- Modular transformation of Jacobi elliptic functions of degree \( n \)

\[
\frac{\text{sn}(z / M; \lambda)}{\text{sn}(z; k)} = \frac{1 - \frac{\text{sn}(z; k)^2}{\text{sn}(2iK'm/ n; k)^2}}{1 - \frac{\text{sn}(z; k)^2}{\text{sn}(2iK'(m - \frac{1}{2})/ n; k)^2}}
\]
Золотарев’s formula: II

\[ \text{sn}(z/k) \]

\[ \text{sn}(z/M, \lambda) \]
Polynomials versus rationals

- Золотарев’s formula has $L_\infty$ error $\Delta \leq e^{\ln \varepsilon}$

- Optimal $L^2$ approximation with weight is

$$\sum_{j=0}^{n} \frac{(-)^j 4}{(2j + 1)\pi} T_{2j+1}(x)$$

  - This has $L^2$ error of $O(1/n)$

- Optimal $L^\infty$ approximation cannot be too much better (or it would lead to a better $L^2$ approximation)
No Free Lunch Theorem

- We must apply the rational approximation with each CG iteration
  - $M^{1/n} \approx r(M)$
  - The condition number for each term in the partial fraction expansion is approximately $\kappa(M)$
  - So the cost of applying $M^{1/n}$ is proportional to $\kappa(M)$
  - Even though the condition number $\kappa(M^{1/n}) = \kappa(M)^{1/n}$
  - And even though $\kappa(r(M)) = \kappa(M)^{1/n}$
- So we don’t win this way…
Instability of symplectic integrator

- Symmetry symplectic integrator
  - Leapfrog
    - Exactly reversible...
    - ...up to rounding errors
- Ляпунов exponent $\nu$
  - $\nu > 0 \ \forall \delta \tau$
    - Chaotic equations of motion (Liu & Jansen)
  - $\nu \propto \delta \tau$ when $\delta \tau$ exceeds critical value $\delta \tau_c$
    - Instability of integrator
  - $\delta \tau_c$ depends on quark mass
Pseudofermions

We want to evaluate a functional integral including the fermionic determinant $\det M$

- We write this as a bosonic functional integral over a pseudofermion field with kernel $M^{-1}$

$$\det M \propto \int d\phi^* d\phi e^{-\phi^* M^{-1} \phi}$$
Multipseudofermions

- We are introducing extra noise into the system by using a single pseudofermion field to sample this functional integral
  - This noise manifests itself as fluctuations in the force exerted by the pseudofermions on the gauge fields
  - This increases the maximum fermion force
  - This triggers the integrator instability
  - This requires decreasing the integration step size

- A better estimate is \( \det M = [\det M^{1/n}]^n \)
  \[
  \det M^{1/n} \propto \int d\phi^* d\phi e^{-\phi^* M^{-1} \phi}
  \]
Hasenbusch’s method

- Clever idea due to Hasenbusch
  - Start with the Wilson fermion action $M = 1 - \kappa H$
  - Introduce the quantity $M' = 1 - \kappa' H$
  - Use the identity $M = M'(M'^{-1}M)$
  - Write the fermion determinant as $\det M = \det M' \det (M'^{-1}M)$
  - Introduce separate pseudofermions for each determinant
  - Adjust $\kappa'$ to minimise the cost

- Easily generalises
  - More than two pseudofermions
  - Wilson-clover action
Violation of NFL Theorem

- So let’s try using our $n^{th}$ root trick to implement multipseudofermions
  - Condition number $\kappa(r(M)) = \kappa(M)^{1/n}$
  - So maximum force is reduced by a factor of $n\kappa(M)^{(1/n)-1}$
  - This is a good approximation if the condition number is dominated by a few isolated tiny eigenvalues

- Cost reduced by a factor of $n\kappa(M)^{(1/n)-1}$
  - Optimal value $n_{opt} \approx \ln \kappa(M)$
  - So optimal cost reduction is $(e \ln \kappa) / \kappa$

- This works!
- Numerical results due to Mike Clark
Acceptance rate at $m=0.025$

- $8^4$ lattice
- $m=0.025$
- $\beta=5.26$
- $\tau=1.0$
- 4 flavours
- naïve staggered
Efficiency at $m=0.025$

- $8^4$ lattice
- $m=0.025$
- $\beta=5.26$
- $\tau=1.0$
- 4 flavours
- naïve staggered
- $n_{opt}=2$
- 33% gain
Acceptance rate at $m=0.01$

- $8^4$ lattice
- $m=0.01$
- $\beta=5.26$
- $\tau=1.0$
- 4 flavours
- naïve staggered
Efficiency at $m=0.01$

- $8^4$ lattice
- $m=0.01$
- $\beta=5.26$
- $\tau=1.0$
- 4 flavours
- naïve staggered
- $n_{opt}=2$
- 60% gain
RHMC technicalities

- In order to keep the algorithm exact to machine precision (as a Markov process)
  - Use a good (machine precision) rational approximation for the pseudofermion heatbath
  - Use a good rational approximation for the HMC acceptance test at the end of each trajectory
  - Use as poor a rational approximation as we can get away with to compute the MD force