Measuring and correcting algorithmic bias in molecular dynamics averages

Stephen Bond

University of Illinois Urbana-Champaign
Department of Computer Science

EPSRC Symposium Workshop on Molecular Dynamics
June 1-5, 2009

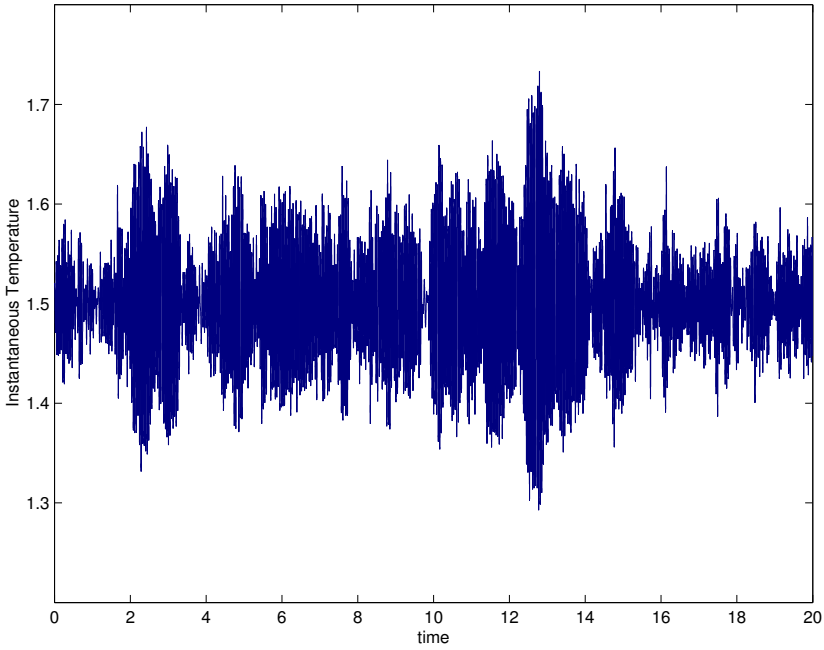
Collaborators

- Nana Arizumi (Illinois)
- Ben Leimkuhler (Edinburgh)
- Brian Laird (Kansas)
- Ruslan Davidchack (Leicester)

S. Bond and B. Leimkuhler *Acta Numerica* **16**, 2007.
N. Arizumi and S. Bond, in preparation, 2009.

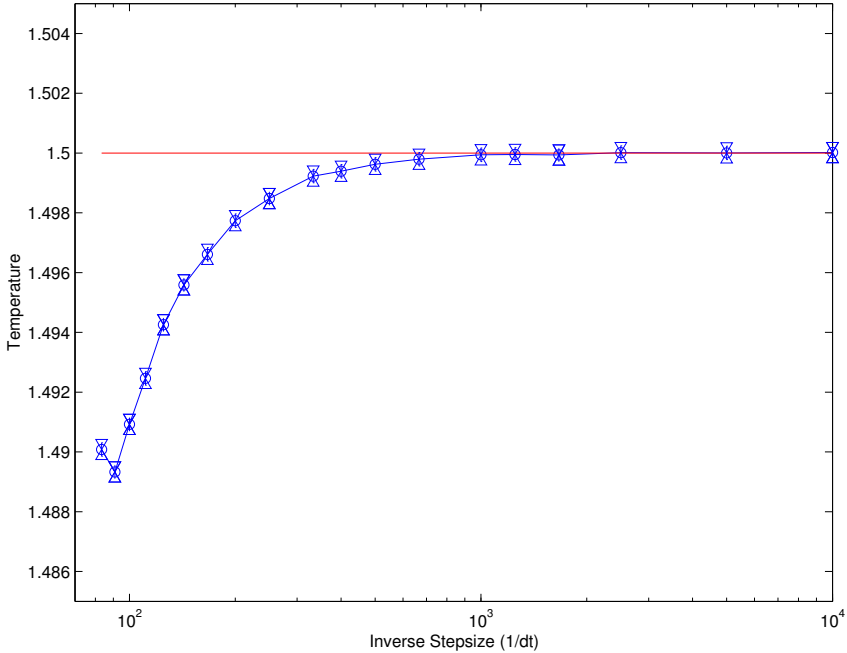
Motivation

Computation of averages



Motivation

Convergence of averages



Motivation

System	Δt [fs]	True temperature [K]	Feedback pressure [bar]	True pressure [bar]
Na^+ in water	1	300.46	1.01	5.14
Na^+ in water	2	301.83	1.24	17.83
TCR-pMHC	1	300.65	1.05	6.97
protein		302.23		
solvent		300.48		

"While the induced temperature error is within one percent, the resulting pressure is manyfold higher than the reference pressure, by a factor of between 5 to 7 with a time step of 1 fs, and a factor of 14 with a time step of 2 fs."

– M.A. Cuendet and W.F. van Gunsteren, *J. Chem. Phys.* **127**, 184102 (2007)

Goal

- What is the error in an average from a MD trajectory?

$$\text{Error} = |\langle A \rangle_{\text{numerical}} - \langle A \rangle_{\text{exact}}|$$

- Estimate accounts for two factors:

$$\text{Error} \leq \text{Statistical Error} + \text{Truncation Error}$$

- Asymptotic Bound:

$$\text{Error} \leq C_1 \frac{1}{\sqrt{t}} + C_2 \, \Delta t^p$$

- Talk will focus on truncation error.

Poincaré hyperbolic: S. Reich, Backward error analysis for numerical integrators, '99

Statistical error: E. Cancès, *et al*, Long-time averaging using symplectic ..., '04, '05.

System of Equations

- Newton's equations: Force = Mass $\times$ acceleration

$$\dot{q} = p/m \quad \text{and} \quad \dot{p} = -\nabla U(q)$$

q = position, m = mass, p = momenta

- First order system

$$\dot{z} = F(z), \quad \text{where} \quad F : \mathbf{R}^n \to \mathbf{R}^n$$

- Exact solution map

$$z(t) = \Phi_t(t_0, z_0)$$

Ergodic

- Time average:

$$\langle A \rangle_{\text{time}} = \lim_{t \to \infty} \frac{1}{t} \int_0^t A\left(z\left(\tau\right)\right) \mathrm{d}\tau$$

- Ensemble average:

$$\langle A \rangle_{\text{ensemble}} = \int_\Omega A\left(z\right) \rho\left(z\right) \mathrm{d}z$$

- Ergodicity

$$\langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}} \qquad (a.e.)$$

Almost all trajectories are statistically the same.

Liouville Equation

- Continuity equation for probability density:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho F) = 0$$

or

$$\frac{\partial \rho}{\partial t} + F \cdot \nabla \rho + \rho \nabla \cdot F = 0$$

- In the case $\rho > 0$,

$$\frac{\mathrm{D} \ln \rho}{\mathrm{D} t} = -\nabla \cdot F$$

- For microcanonical ensemble

$$\nabla \cdot F = 0 \quad \Rightarrow \quad \rho = C \; \delta \left[H(z) - E \right]$$

Example: Nosé-Hoover

- Nosé-Hoover vector field

$$\begin{aligned} \frac{dq}{dt} &= M^{-1}p \\ \frac{dp}{dt} &= -\nabla U(q) - \frac{\xi}{\mu}p \\ \frac{d\xi}{dt} &= p^T M^{-1} p - g k_B T \end{aligned}$$

- Invariant distribution

$$\rho \propto \exp\left\{ -\frac{1}{k_B T}\left(\frac{1}{2} p^T M^{-1} p + U(q) + \frac{\xi^2}{2\mu} \right) \right\}$$

Error Analysis

- First order system
$$\dot{z} = F(z)$$
- Forward error:
Is the numerical trajectory close to the exact trajectory?
$$\| z_n - z(t_n) \| \leq C \Delta t^p$$
- Backward error:
Is the numerical trajectory interpolated by an exact trajectory, but for a different problem?
$$\| \hat{F}_{\Delta t}(z) - F(z) \| \leq C \Delta t^p$$

"Method of Modified Equations"

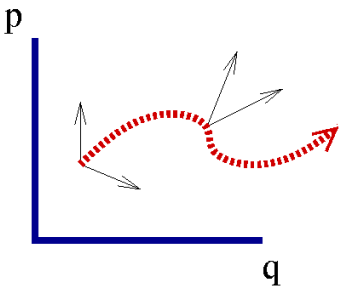
Error Analysis

- Ergodicity:
  Exact trajectories are sensitive (chaotic) to perturbations in the initial conditions
  $\rightarrow$ Large Forward Error.

- Statistics:
  Thermodynamic properties (averages) are not a function of the details of the initial conditions
  $\rightarrow$ Small Backward Error.

Symplectic Structure

- In 1 dimension



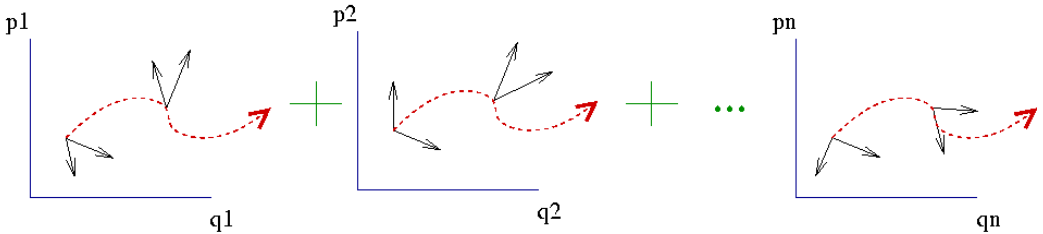
Conservation of Area

$dq \wedge dp = \text{constant}$

- In n dimensions

Conservation of "Oriented Area"

$\sum dq_i \wedge dp_i = \text{constant}$



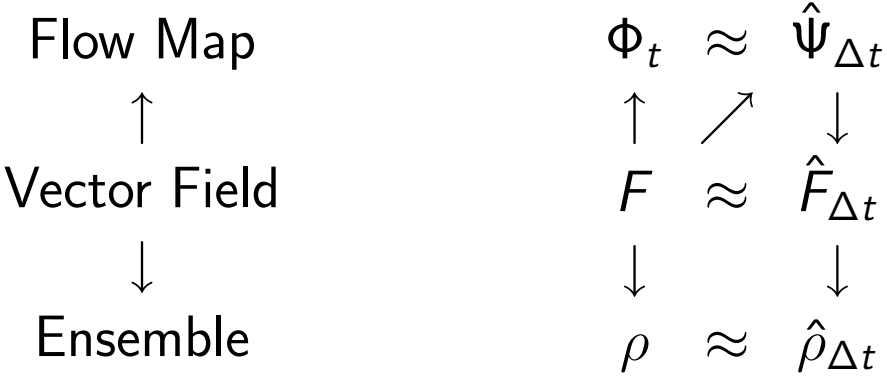
Backward Error Analysis: Modified Equations

- Given a pth-order numerical method, Ψ_h, we can always construct a modified vector field, F_h, such that the numerical method provides a rth-order approximation to the flow of the modified system .
- If the numerical method and vector field are symplectic/Hamiltonian, the modified vector field will be symplectic/Hamiltonian.

Backward Error Analysis: Modified Equations

- Given a pth-order numerical method, Ψ_h, we can always construct a modified vector field, F_h, such that the numerical method provides a rth-order approximation to the flow of the modified system .
- If the numerical method and vector field are symplectic/Hamiltonian, the modified vector field will be symplectic/Hamiltonian.
- Series is truncated at an optimal r^*, which increases as $h \longrightarrow 0$.
- Use a low-order modified vector field when h is large?

Big Picture


Flow Map
Vector Field
Ensemble
$\Phi_t \approx \hat{\Psi}_{\Delta t}$
$F \approx \hat{F}_{\Delta t}$
$\rho \approx \hat{\rho}_{\Delta t}$

Example: Verlet

- Hamiltonian

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

- Verlet

$$\begin{aligned} p^{n+1/2} &= p^n - \frac{\Delta t}{2} \nabla U(q^n) \\ q^{n+1} &= q^n + \Delta t M^{-1} p^{n+1/2} \\ p^{n+1} &= p^{n+1/2} - \frac{\Delta t}{2} \nabla U\left(q^{n+1}\right) \end{aligned}$$

- Splitting

$$H_1 = \frac{1}{2} p^T M^{-1} p, \qquad H_2 = U(q)$$

Example: Verlet

- Strang Splitting

$$\exp\left(\Delta t\mathcal{L}\right) = \exp\left(\frac{\Delta t}{2}\mathcal{L}_2\right)\exp\left(\Delta t\mathcal{L}_1\right)\exp\left(\frac{\Delta t}{2}\mathcal{L}_2\right) + \mathcal{O}\left[\Delta t^3\right]$$

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$$

$$\mathcal{L}_1 = M^{-1}p\cdot\nabla_q \qquad \mathcal{L}_2 = -\nabla_q U\left(q\right)\cdot\nabla_p$$

- Modified Equations

$$\exp\left(\Delta t\hat{\mathcal{L}}^{[r]}_{\Delta t}\right) = \exp\left(\frac{\Delta t}{2}\mathcal{L}_2\right)\exp\left(\Delta t\mathcal{L}_1\right)\exp\left(\frac{\Delta t}{2}\mathcal{L}_2\right) + \mathcal{O}\left[\Delta t^{r+1}\right]$$

Solve for $\hat{\mathcal{L}}^{[r]}_{\Delta t}$ using Baker-Campbell-Hausdorff formula

Example: Verlet

- Original Hamiltonian:

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

- Modified Hamiltonian:

$$H_{2,\Delta t}(q,p) = H(q,p) + \frac{\Delta t^2}{12}\left(p^T M^{-1} U'' M^{-1} p - \frac{1}{2}\nabla U^T M^{-1} \nabla U\right)$$

Verlet conserves $H_{2,\Delta t}$ to 4th order accuracy!

Example: Verlet

- Original Hamiltonian:

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

- Modified Hamiltonian:

$$H_{2,\Delta t}(q,p) = H(q,p) + \frac{\Delta t^2}{12}\left(p^T M^{-1} U'' M^{-1} p - \frac{1}{2} \nabla U^T M^{-1} \nabla U \right)$$

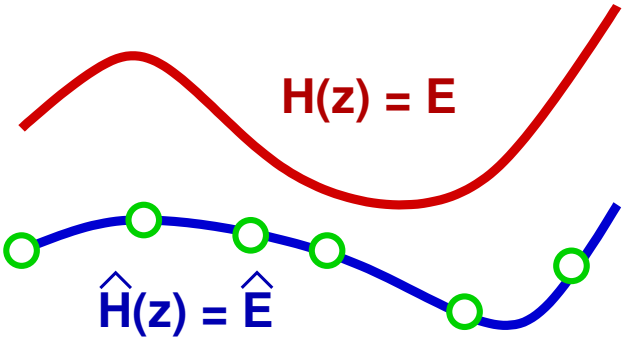
Verlet conserves $H_{2,\Delta t}$ to 4th order accuracy!

- Practical Computation:

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}t^2} U(q) &= \frac{\mathrm{d}}{\mathrm{d}t} \nabla U(q) \cdot M^{-1} p \\ &= p^t M^{-1} U''(q) M^{-1} p - \nabla U(q) \cdot M^{-1} \nabla U(q) \end{aligned}$$

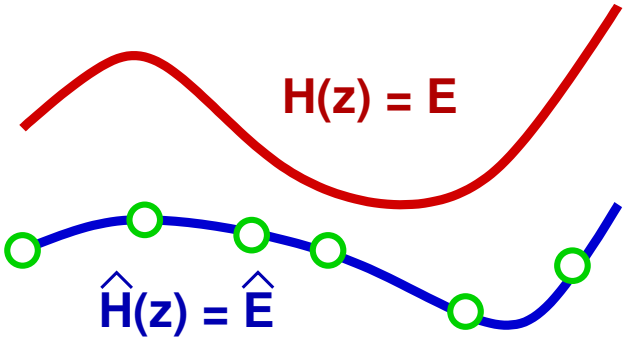
Correcting Microcanonical Averages

- Hamiltonian + Symplectic integrator $\Rightarrow$ Modified Hamiltonian



- Numerical average is computed on $\hat{H}$ surface
- What is the error from using the wrong surface?

Correcting Microcanonical Averages



$$\text{Exact} - \text{Numerical} \approx \langle A \rangle_{H=E} - \langle A \rangle_{\hat{H}=\hat{E}} =$$

$$\frac{\int A(z)\, \delta\left[H(z) - E\right] \mathrm{d}z}{\int \delta\left[H(z) - E\right] \mathrm{d}z} - - \frac{\int A(z)\, \delta\left[\hat{H}(z) - \hat{E}\right] \mathrm{d}z}{\int \delta\left[\hat{H}(z) - \hat{E}\right] \mathrm{d}z}$$

Correcting Microcanonical Averages

- Expand delta function

$$\delta\Big[H(z)\Big] = \delta\Big[\hat{H}(z)\Big] + \Big(H - \hat{H}\Big)\,\delta'\Big[\hat{H}(z)\Big] + \cdots$$

and use directional derivative

$$u \cdot \nabla_z\, \delta\Big[H(z)\Big] = \delta'\Big[H(z)\Big]\, u \cdot \nabla H$$

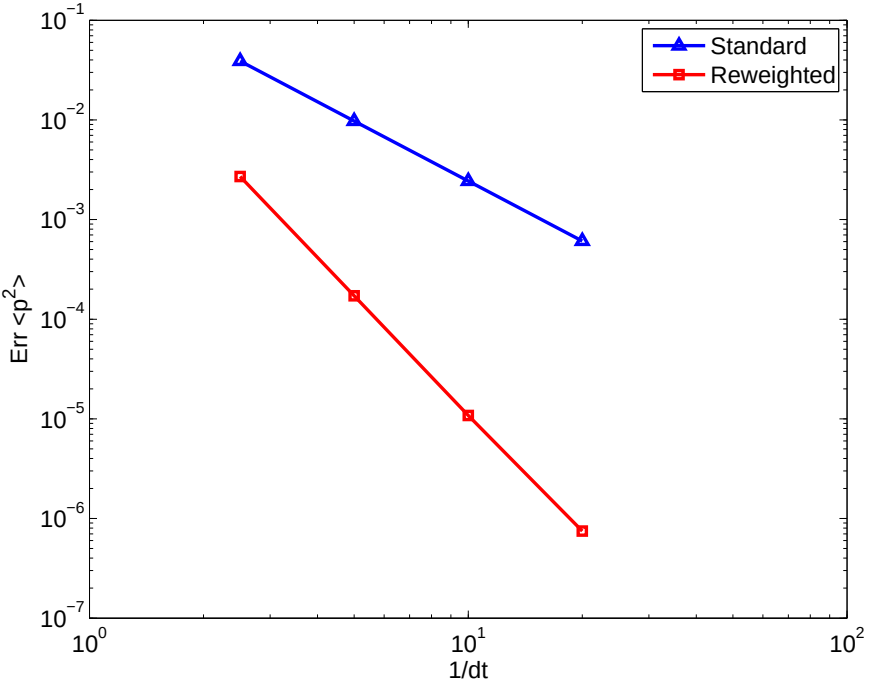
- Corrected average

$$\langle A \rangle_{\text{exact}} = \frac{\langle A \rangle_{\text{num}} + \langle \nabla \cdot (w\,A) \rangle_{\text{num}}}{\langle 1 \rangle_{\text{num}} + \langle \nabla \cdot w \rangle_{\text{num}}} + \cdots$$

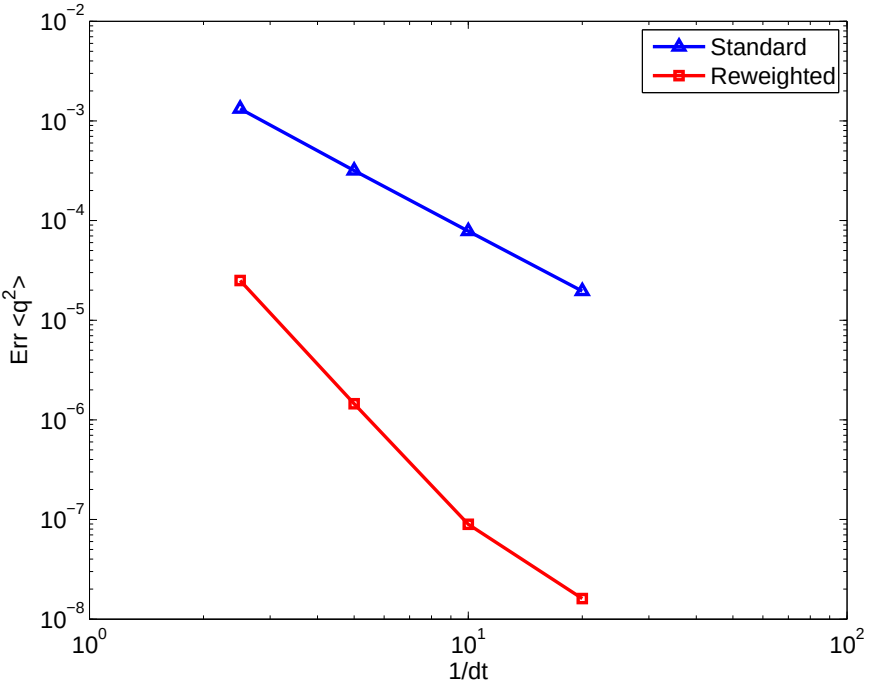
where

$$w := \Big(\hat{H} - H\Big)\,\frac{u}{u \cdot \nabla \hat{H}}$$

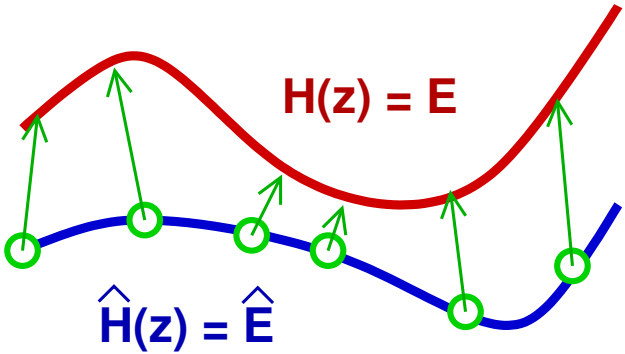
Example: Quartic Oscillator



Example: Quartic Oscillator



Correcting Microcanonical Averages



- Alternative method:

$$\langle A \rangle_{\text{exact}} = \langle \omega \; A(T(z)) \rangle_{\text{num}}$$

where T maps points on $\hat{H}$ to points on H.

- Weighting factor, ω, accounts for distortion

Liouville Equation for Modified Vector Field

- Modified Equations

$$\frac{d\hat{z}}{dt} = \hat{F}_{\Delta t}(\hat{z}) \qquad \text{where} \quad \hat{F}_{\Delta t} = F + \Delta t^p G$$

- Modified Liouville Equation

$$\frac{\partial}{\partial t}\hat{\rho}_{\Delta t} + \nabla \cdot \left(\hat{\rho}_{\Delta t}\, \hat{F}_{\Delta t}\right) = 0$$

- Weighting factor

$$\omega := \rho / \hat{\rho}_{\Delta t}, \quad \text{assuming} \quad \rho, \hat{\rho}_{\Delta t} > 0$$

implies

$$\frac{\overline{\mathrm{D}}}{\overline{\mathrm{D}}t} \ln(\omega) = \Delta t^p \left(\nabla \cdot G + G \cdot \nabla \ln \rho\right)$$

Averages

- Truncation Error Estimate

$$\begin{aligned} \langle A \rangle_{\text{num}} - \langle A \rangle_{\text{exact}} &\approx \int_{\Gamma} A(q,p)\, \rho_{\Delta t}\, \mathrm{d}\Gamma - \int_{\Gamma} A(q,p)\, \rho\, \mathrm{d}\Gamma \\ &\approx \frac{\langle A \rangle_{\text{num}} \langle \omega \rangle_{\text{num}} - \langle A\, \omega \rangle_{\text{num}}}{\langle \omega \rangle_{\text{num}}} \end{aligned}$$

- Reweighted Averages

$$\langle A \rangle_{\text{exact}} = \frac{\langle A\, \omega \rangle_{\text{num}}}{\langle \omega \rangle_{\text{num}}} + \mathcal{O}\left[\Delta t^{r}\right]$$

Example:

- Nosé-Poincaré Hamiltonian:

$$H(q, \tilde{p}, s, \pi_s) = s\left(\frac{1}{2\,s^2}\tilde{p}^T M^{-1}\tilde{p} + U(q) + \frac{\pi_s^2}{2\mu} + g\,k\,T\,\ln s - E_0\right)$$

- Nosé-Poincaré Modified Hamiltonian:

$$\begin{aligned}\hat{H}_{\Delta t} &= H_{NP} + \frac{\Delta t^2}{12} s\left(\frac{\pi_s}{\mu\, s}\tilde{p}^T M^{-1}\nabla U\right. \\ &- \frac{1}{2}\nabla U^T M^{-1}\nabla U + \frac{1}{s^2}\tilde{p}^T M^{-1} U'' M^{-1}\tilde{p} \\ &- \left.\frac{1}{2\,\mu}\left(\frac{1}{s^2}\tilde{p}^T M^{-1}\tilde{p} - g\,k\,T\right)^2 + \frac{2\,g\,k\,T\,\pi_s^2}{\mu^2}\right)\end{aligned}$$

Example:

- Modified marginal distribution:

$$\begin{aligned}\bar{\rho}_{\Delta t}(q,p)\,\mathrm{d}p\,\mathrm{d}q &= \frac{1}{C}\int\!\!\int_{p_s}\delta\left[\hat{H}_{\Delta t}(q,s,\tilde{p},p_s)-\hat{E}_0\right]\mathrm{d}\tilde{p}\,\mathrm{d}q\,\mathrm{d}p_s\,\mathrm{d}s,\\ &= \frac{1}{C}\int\!\!\int_{p_s}\delta\left[s\left(H_{\mathrm{N}}-H_{\mathrm{N}}^0+\Delta t^2 G\right)\right]\mathrm{d}\tilde{p}\,\mathrm{d}q\,\mathrm{d}p_s\,\mathrm{d}s.\end{aligned}$$

- Change of variables, integrating

$$\bar{\rho}=\frac{1}{C}\int_{p_s}\mathrm{e}^{N_f\eta_0}\left|gk_BT+h^2\frac{\partial}{\partial\eta}G(q,\mathrm{e}^{\eta},p,p_s)\right|_{\eta=\eta_0}^{-1}\mathrm{d}p_s.$$

$$\eta_0=\frac{-1}{g\,k_B\,T}\left(H(q,p)+\frac{p_s^2}{2\,\mu}+h^2\,G(q,\mathrm{e}^{\eta_0},p,p_s)-H_{\mathrm{N}}^0\right),$$

- More mathematical manipulations

$$\bar{\rho}=\frac{\rho_c}{\bar{C}}\exp\left\{-\frac{\Delta t^2}{24k_BT}\left[\sum_{j,k}\frac{2p_jp_kU_{q_jq_k}}{m_jm_k}-\sum_j\frac{U_{q_j}^2}{m_j}-\frac{1}{\mu}\left(\sum_j\frac{p_j^2}{m_j}-gk_BT\right)^2\right]\right\},$$

Example:

- Weighting Factor:

$$\omega \approx \exp\left\{\frac{-\Delta t^2}{24\, k_{\mathrm{B}}\, T}\left[2p^T M^{-1} U''(q)\, M^{-1} p - \nabla U(q)^T\, M^{-1} \nabla U(q) - \frac{1}{\mu}\left(p^T M^{-1} p - g\, k_{\mathrm{B}}\, T\right)^2\right]\right\}$$

- Reweighted Averages:

$$\langle A \rangle_{\mathrm{exact}} \approx \frac{\langle A\, \omega \rangle_{\mathrm{num}}}{\langle \omega \rangle_{\mathrm{Num}}}$$

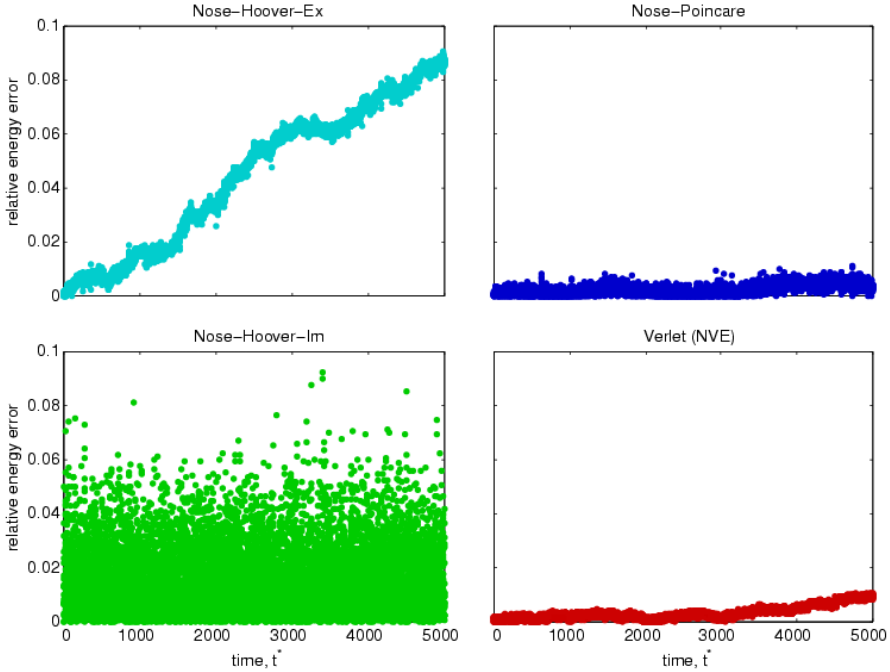
- Hybrid Monte Carlo:
  J. Izaguirre and S. Hampton, *J. Comput. Phys.* **200**, 2004.
  E. Akhmatskaya and S. Reich, *LNCSE* **49**, 2006.
- Time correlation functions:
  R. D. Skeel, *SIAM J. Sci. Comput.* **31**, 2009.

Numerical Experiment:

- System:
  - 256 Particle Gas
  - Lennard-Jones Potential
  - $T = 1.5\epsilon/k$, $\rho = 0.95 r_0^3$, $t = 20 r_0 \sqrt{m/\epsilon}$
- Method:
  - Nosé-Poincaré (Symplectic, Time-Reversible)
  - $\Delta t = 0.012 r_0 \sqrt{m/\epsilon}$ to $0.0001 r_0 \sqrt{m/\epsilon}$
- Reference:
  - Dettmann and Morriss, *Phys. Rev. E* **55** 1997
  - Bond, Laird, and Leimkuhler *J. Comput. Phys.* **151** 1999.
  - S. Bond and B. Leimkuhler *Acta Numerica* **16**, 2007.

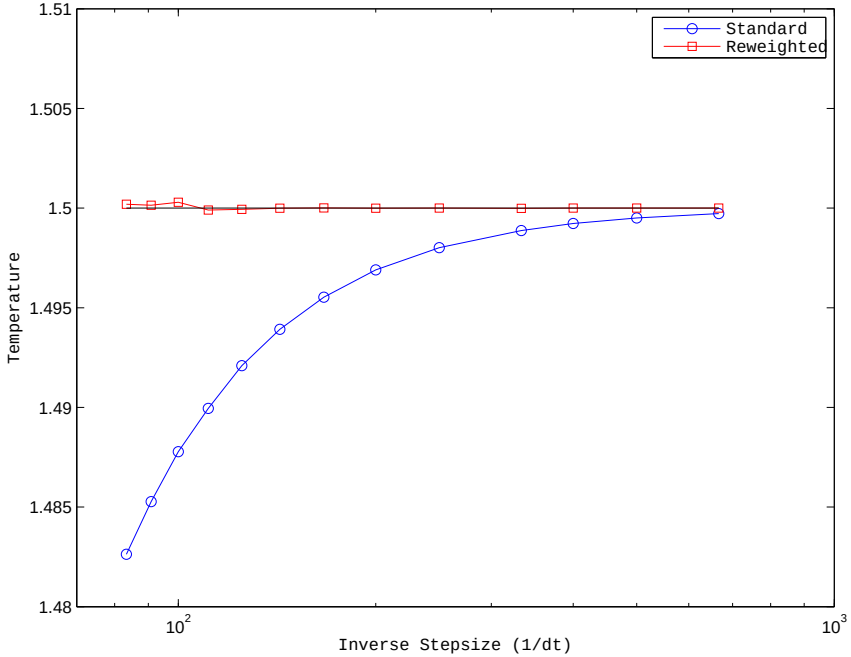
Numerical Experiment:

- "Extended" Energy Conservation:



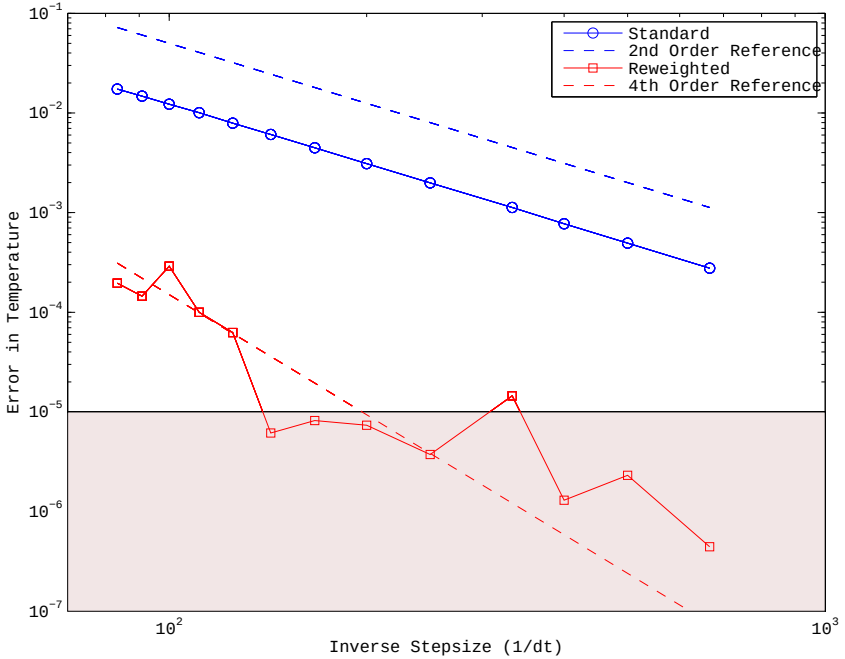
Numerical Experiment:

- Improved Estimator



Numerical Experiment:

- Improved Estimator Error



Final Thoughts:

- Numerical stability of pressure measurement
- In general, when can we correct for numerical bias?
- Is the error in the dynamics or the observation?

Final Thoughts:

- Numerical stability of pressure measurement
- In general, when can we correct for numerical bias?
- Is the error in the dynamics or the observation?
- Ruslan Davidchack, Warwick Capstone Minisymposium (July 1, 2009)

Goal-Oriented Error Estimation and Multilevel Preconditioning for the Poisson-Boltzmann Equation

Stephen Bond

University of Illinois at Urbana Champaign
Department of Computer Science

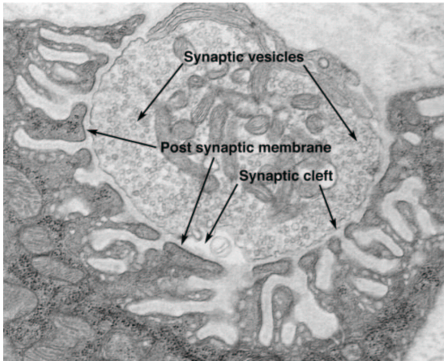
June 1-5, 2009

Acknowledgements

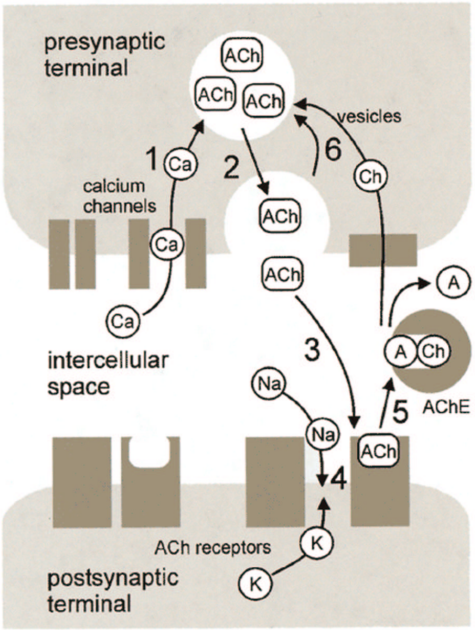
- Eric Cyr (Illinois / Sandia)
- Michael Holst (UCSD)
- Andrew McCammon (UCSD)
- Burak Aksoylu (LSU)
- Nathan Baker (Wash U)
- Kaihsu Tai (Oxford)
- Hugh MacMillan (Clemson)

Motivation

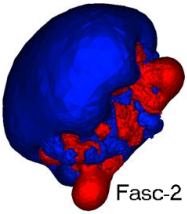
Green Mamba: Fasciculin 2

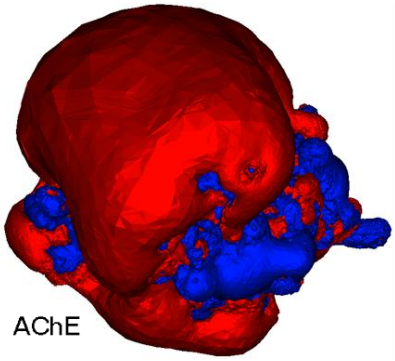


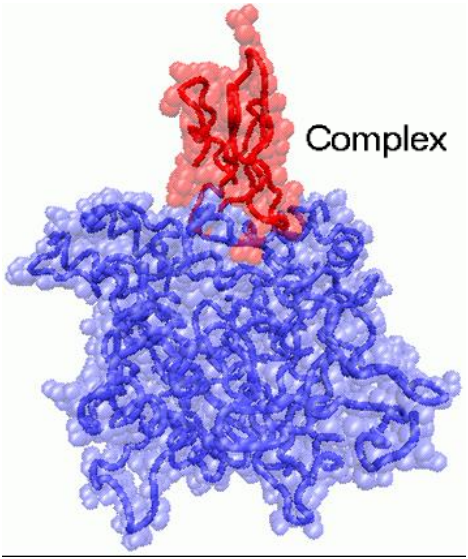
Neuromuscular Junction



Motivation







Outline

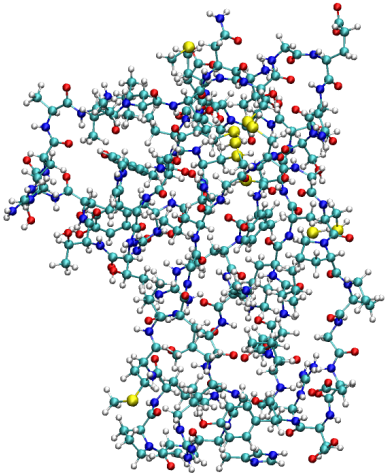
Motivation

Proteins naturally occur in solution

$\Rightarrow$ Must model them in solution

Two options for modeling solute-solvent electrostatic interactions

1. *Explicit*: Solvent molecules explicitly represented
2. *Implicit*: "Average" effect of solvent is computed



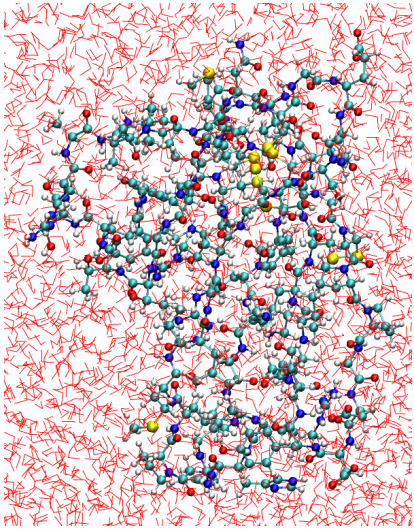
Motivation

Proteins naturally occur in solution

$\Rightarrow$ Must model them in solution

Two options for modeling solute-solvent electrostatic interactions

1. *Explicit*: Solvent molecules explicitly represented
2. *Implicit*: "Average" effect of solvent is computed



Motivation

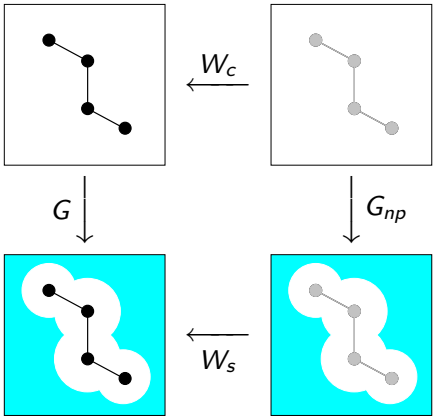
Want to compute electrostatic Solvation Free Energy:

- Total Solvation Free Energy

$$G = W_s - W_c + G_{np}$$

- Electrostatic Solvation Free Energy

$$S = W_s - W_c$$

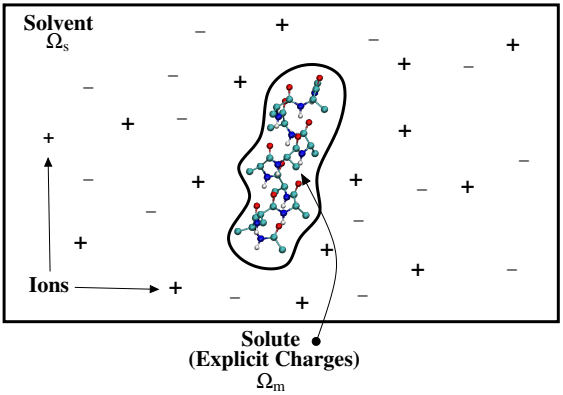


Outline

Formulation

Poisson-Boltzmann Equation (PBE): Nonlinear PDE to compute electrostatics of protein in solution

- 3D infinite domain: Ω
- 2 Subdomains:
  - Solute: Ω_m
  - Solvent: Ω_s
- Interface: Γ



Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x) \nabla \phi(x) + \bar{\kappa}^2(x) \sinh\left(\phi(x)\right) &= 4\pi \rho_f(x) \quad \text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) &= 0 \quad \text{for} \quad x = \infty, \\
[\epsilon(x) \nabla \phi(x) \cdot n] &= 0 \quad \text{for} \quad x \in \Gamma.
\end{aligned}
$$

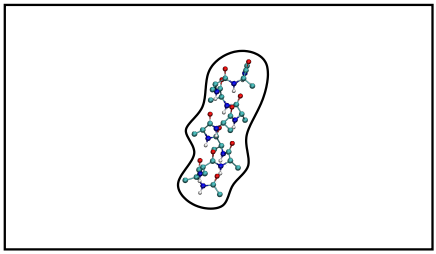
Note: ϵ and $\bar{\kappa}$ are discontinuous at interface
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain



Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x) \nabla \phi(x) + \bar{\kappa}^2(x) \sinh\left(\phi(x)\right) &= 4\pi \rho_f(x) \quad \text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) &= 0 \quad \text{for} \quad x = \infty, \\
[\epsilon(x) \nabla \phi(x) \cdot n] &= 0 \quad \text{for} \quad x \in \Gamma.
\end{aligned}
$$

Note: ϵ and $\bar{\kappa}$ are discontinuous at interface

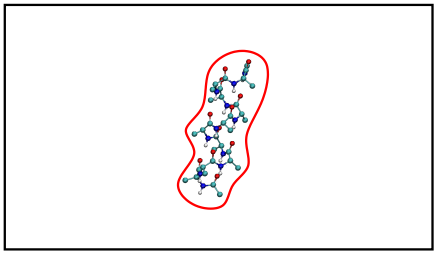
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain



Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x)\nabla\phi(x) + \bar{\kappa}^2(x)\sinh\left(\phi(x)\right) = 4\pi\rho_f(x) \quad &\text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) = g(x) \quad &\text{for} \quad x \in \partial\Omega, \\
[\epsilon(x)\nabla\phi(x) \cdot n] = 0 \quad &\text{for} \quad x \in \Gamma.
\end{aligned}
$$

Note: ϵ and $\bar{\kappa}$ are discontinuous at interface
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain

Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x)\nabla\phi(x) + \bar{\kappa}^2(x)\sinh\left(\phi(x)\right) &= 4\pi\rho_f(x) \quad \text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) &= g(x) \quad \text{for} \quad x \in \partial\Omega, \\
[\epsilon(x)\nabla\phi(x) \cdot n] &= 0 \quad \text{for} \quad x \in \Gamma.
\end{aligned}
$$

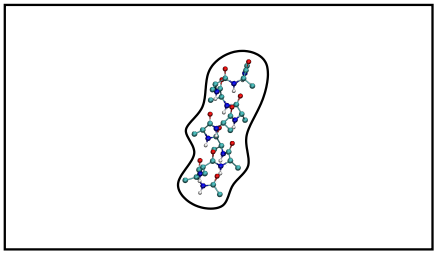
Note: ϵ and $\bar{\kappa}$ are discontinuous at interface
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain



Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x) \nabla \phi(x) + \bar{\kappa}^2(x) \phi(x) &= 4\pi \rho_f(x) \quad \text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) &= g(x) \quad \text{for} \quad x \in \partial\Omega, \\
[\epsilon(x) \nabla \phi(x) \cdot n] &= 0 \quad \text{for} \quad x \in \Gamma.
\end{aligned}
$$

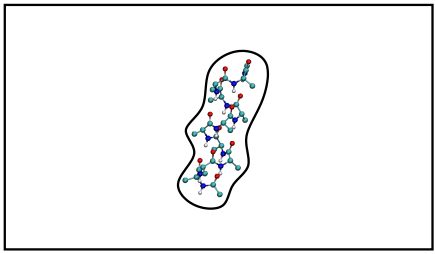
Note: ϵ and $\bar{\kappa}$ are discontinuous at interface
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain



Formulation: PDE

Poisson-Boltzmann Equation (PBE) for electrostatic potential

$$
\begin{aligned}
-\nabla \cdot \epsilon(x)\nabla\phi(x) + \bar{\kappa}^2(x)\phi(x) &= 4\pi\rho_f(x) \quad \text{for} \quad x \in \Omega_m \cup \Omega_s, \\
\phi(x) &= g(x) \quad \text{for} \quad x \in \partial\Omega, \\
[\epsilon(x)\nabla\phi(x) \cdot n] &= 0 \quad \text{for} \quad x \in \Gamma.
\end{aligned}
$$

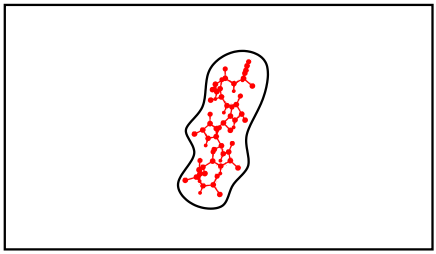
Note: ϵ and $\bar{\kappa}$ are discontinuous at interface
Simplifications:

- Infinite domain assumed to be finite
- Linearized PBE (LPBE) assumes $\phi \approx \sinh(\phi)$

Challenge:

- ρ_f: point charges at solute atom locations cause singularities in ϕ

PBE Domain



Formulation: What Challenge?

Source term: $\rho_f(x) = \sum_i^P q_i \delta(x - x_i)$

Coulomb's law for single charge is (ϵ_m is dielectric in vacuum)

$$-\epsilon_m \nabla^2 G(x) = 4\pi\delta(x) \quad \Rightarrow \quad G(x) = \frac{1}{\epsilon_m |x|}$$

Consider PBE in molecular subdomain Ω_m (i.e. $\bar{\kappa}^2(x) = 0$)

$$-\epsilon_m \nabla^2 \phi(x) = 4\pi\rho_f(x)$$

- Up to $\mathrm{Ker}(\nabla^2)$, ϕ is given by Coulomb's law
- Standard piecewise linear FE basis does not converge to $1/|x|$ singularities

Formulation: Regularized PBE

Will remove singularities from electrostatic potential analytically

- Use analytical form of Coulomb potential $G(x)$, satisfying

$$-\epsilon_m \nabla^2 G(x) = 4\pi \rho_f(x)$$
$$G(\infty) = 0$$

- Define $u = \phi - G$, u called Reaction Potential
- Substitute $\phi = u + G$ into PBE, solve for u gives Regularized PBE (RPBE)

$$\left.\begin{aligned} -\nabla \cdot \epsilon(x) \nabla u(x) + \bar{\kappa}^2(x) u(x) \\ = \nabla \cdot (\epsilon(x) - \epsilon_m) \nabla G - \bar{\kappa}^2(x) G(x) \end{aligned}\right\} \quad \text{for} \quad x \in \Omega_m \cup \Omega_s,$$

$$u(x) = g(x) - G(x) \quad \text{for} \quad x \in \partial\Omega,$$

$$[\epsilon(x) \nabla u(x) \cdot n] = (\epsilon_m - \epsilon_s) \nabla G(x) \cdot n \quad \text{for} \quad x \in \Gamma.$$

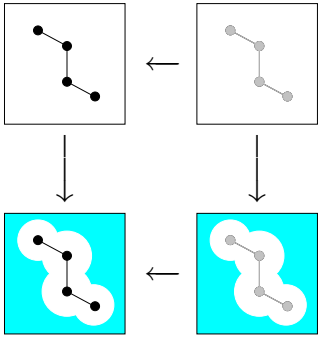
Goto weak form

Formulation: Solvation Free Energy

Recall, goal is to compute Solvation Free Energy: $S = W_s - W_c$

- $u = \phi - G$
- S is a linear functional of u

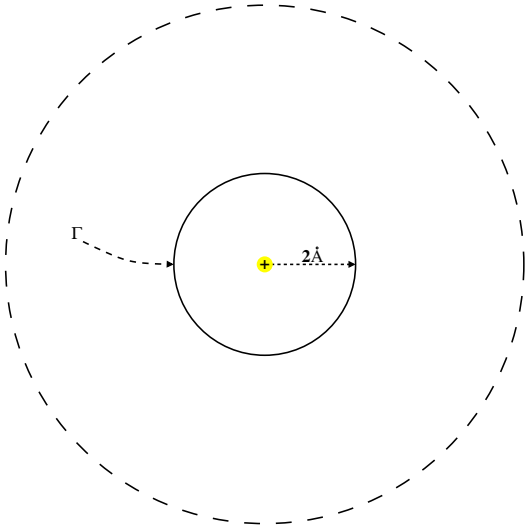
$$S(u) = \frac{1}{2} \int u(x) \rho_f(x) \, dx$$

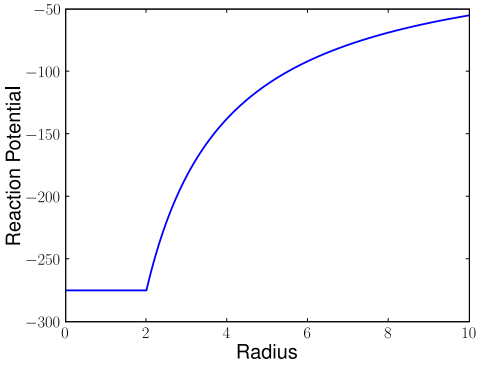


Formulation: Born ion example

Born ion is single ion in center of sphere: analytical solution known

$\Rightarrow$ Sphere radius$= 2\text{Å}$, $\epsilon_m = 1$, $\epsilon_s = 78$, $\bar{\kappa}^2 = 0$



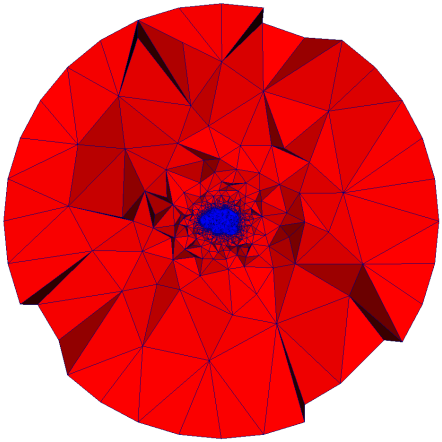


Outline

Discretization

Discretization concerns

- Complex geometry of molecule $\Rightarrow$ Use Finite Elements
- Problem: Need mesh matching molecular surface
  Use GAMer (Geometry preserving Adaptive MeshER) from Holst group *



*Z. Yu, M. Holst, Y. Cheng, and J.A. McCammon, J. Mol. Graphics, 2008.

Discretization

Use finite elements to solve linear RPBE

- Primal problem is

$$a(u, v) = L(v) \quad \forall v \in V$$

where

$$a(u, v) = \int_\Omega \epsilon(x) \nabla u(x) \cdot \nabla v(x) + \bar{\kappa}^2(x) u(x) v(x) \, dx$$

$$L(v) = \int_\Omega -(\epsilon(x) - \epsilon_m) \nabla G(x) \cdot \nabla v(x) - \bar{\kappa}^2 G(x) v(x) \, dx.$$

- Want to compute solvation free energy: $S(u)$

Initial mesh not good enough for accurate solvation free energy

- Use Adaptive Mesh Refinement to improve accuracy

see RPBE

Outline

Adaptive Mesh Refinement

Adaptive Mesh Refinement (AMR) Algorithm

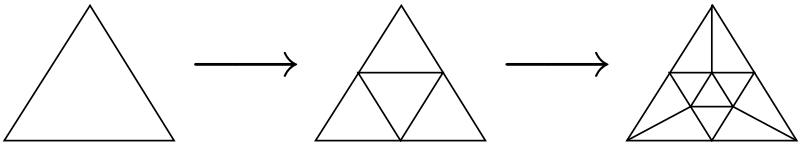
SOLVE $\longrightarrow$ ESTIMATE $\longrightarrow$ MARK $\longrightarrow$ REFINE

SOLVE: Solve Regularized LPBE

ESTIMATE: Construct elementwise error estimates

MARK: Select elements with "large" error for refinement

REFINE: Subdivide selected elements into smaller simplices



Adaptive Mesh Refinement

Adaptive Mesh Refinement (AMR) Algorithm

SOLVE $\longrightarrow$ ESTIMATE $\longrightarrow$ MARK $\longrightarrow$ REFINE

SOLVE: Solve Regularized LPBE

ESTIMATE: Construct elementwise error estimates

MARK: Select elements with "large" error for refinement

REFINE: Subdivide selected elements into smaller simplices

I will focus on ESTIMATE and MARK

AMR - ESTIMATE

From approximate solution u^h calculate error

1. Easily computed
2. Bounds the error
   - What error should be bounded?

How to compute error in u^h?

- Relate weak residual to error

$$R(v) = L(v) - a(u^h, v) \quad \forall v \in V$$

AMR - ESTIMATE: Energy-based

Measure error in energy-norm ($\|g\|^2 = a(g,g)$)

$$\left\| u - u^h \right\|^2 \leq C \left(\sum_K \eta_K^2(u^h) \right)$$

Indicator is

$$\eta_K^2(u^h) = h_K^2 \|r_K\|_{L^2(K)}^2 + \frac{1}{4} h_{\partial K} \|r_{\partial K}\|_{L^2(\partial K)}^2$$

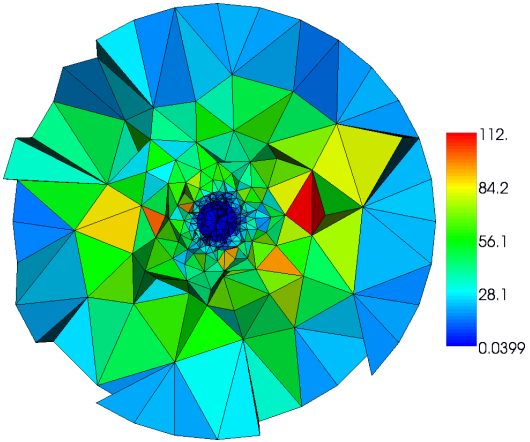
where

$$r_K(x) = (\nabla \cdot (\epsilon(x) - \epsilon_m) \nabla G(x) - \bar{\kappa}^2(x) G(x)) - (-\nabla \cdot \epsilon(x) \nabla u^h(x) + \bar{\kappa}^2(x) u^h(x))$$
$$r_{\partial K}(x) = n_K \cdot \left[(\epsilon(x) - \epsilon_m) \nabla G(x) + \epsilon(x) \nabla u^h(x) \right]_{n_K}$$

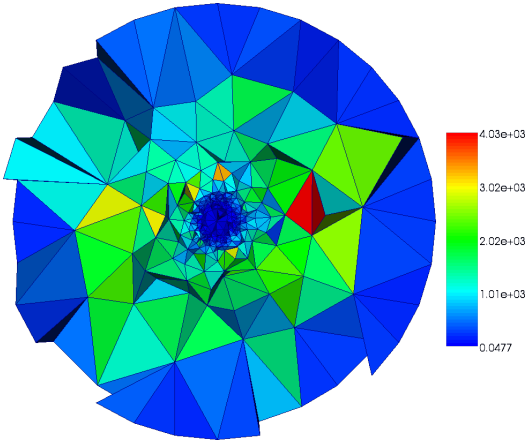
AMR - ESTIMATE: Energy-based

Elementwise error in solution of RPBE for Born ion

- ✓ Indicator shows correct error distribution
- × Scaling of indicator is wrong
- × No explicit knowledge of Solvation Free Energy



Exact Error



Numerical Indicator

AMR - ESTIMATE: Goal-oriented

Want to find error in solvation free energy: $S(u - u^h)$

- By Riesz Representation there exits a w such that

$$R(w) = a(u - u^h, w) = S(u - u^h)$$

- Given w, error in $S(u^h)$ can be computed
- To find w, solve the Dual problem

$$a(v, w) = S(v) \quad \forall v \in V$$

- In practice, w is approximated by FE solution
- Indicators bound error in functional

$$|S(u - u^h)| = |a(u - u^h, w)| \leq C \sum_K \eta_K(u^h, w^h)$$

AMR - ESTIMATE: Goal-oriented

Algorithm: Goal-Oriented Refinement

1. Solve primal problem for u^h
2. Solve dual problem for w^h
3. Compute error indicator

$$|S(u - u^h)| = |a(u - u^h, w)| \leq \sum_K \eta_K(u^h, w^h)$$

   where K is an element
4. Refine elements where $\eta_K(u^h, w^h)$ is "large"
5. Repeat

AMR - ESTIMATE: Goal-oriented

Two options for computing η_K

1. Solve dual problem using quadratic basis functions: $w \approx w^{h,2}$

$$|S(u-u^h)| = |L(w^{h,2}) - a(u^h, w^{h,2})| \leq \sum_K \eta_K(u^h, w^{h,2})$$

where

$$\eta_K(u^h, w^{h,2}) = \int_K \Big| (\epsilon - \epsilon_m)\nabla G(x)\nabla w^{h,2}(x) + \bar{\kappa}^2(x)G(x)w^{h,2}(x) + \epsilon(x)\nabla u^h(x)\cdot\nabla w^{h,2}(x) + \bar{\kappa}^2(x)u^h(x)w^{h,2}(x)\Big| dx$$

2. Solve dual problem using linear basis functions

$$|S(u-u^h)| = \sum_K \frac{1}{4}\|(u-u^h)+(w-w^h)\|_K^2 - \frac{1}{4}\|(u-u^h)-(w-w^h)\|_K^2$$

where $u-u^h$ and $w-w^h$ are approximated using element residual method

AMR - MARK

MARK: Select elements with "large" error for refinement

- Choice of marking greatly effects quality of refinement

For a triangulation $\mathcal{T} = \mathcal{T}^m \cup \mathcal{T}^s$: two marking strategies

1. Global Marking: For $\gamma \in (0, 1)$

$$\text{Mark all } K \in \mathcal{T} \text{ such that } \eta_K > \gamma \max_{T \in \mathcal{T}} \eta_T$$

2. Split Marking: For $\gamma \in (0, 1)$

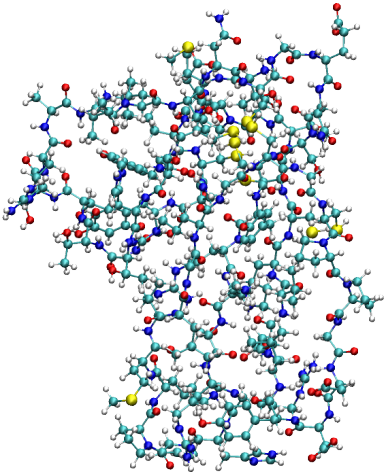
$$\text{Mark all } \begin{cases} K \in \mathcal{T}^s \\ K \in \mathcal{T}^m \end{cases} \text{ such that } \begin{cases} \eta_K > \gamma \max\limits_{T \in \mathcal{T}^s} \eta_T \\ \eta_K > \gamma \max\limits_{T \in \mathcal{T}^m} \eta_T \end{cases}$$

Implicit Solvent Results

Compute solvation free energy

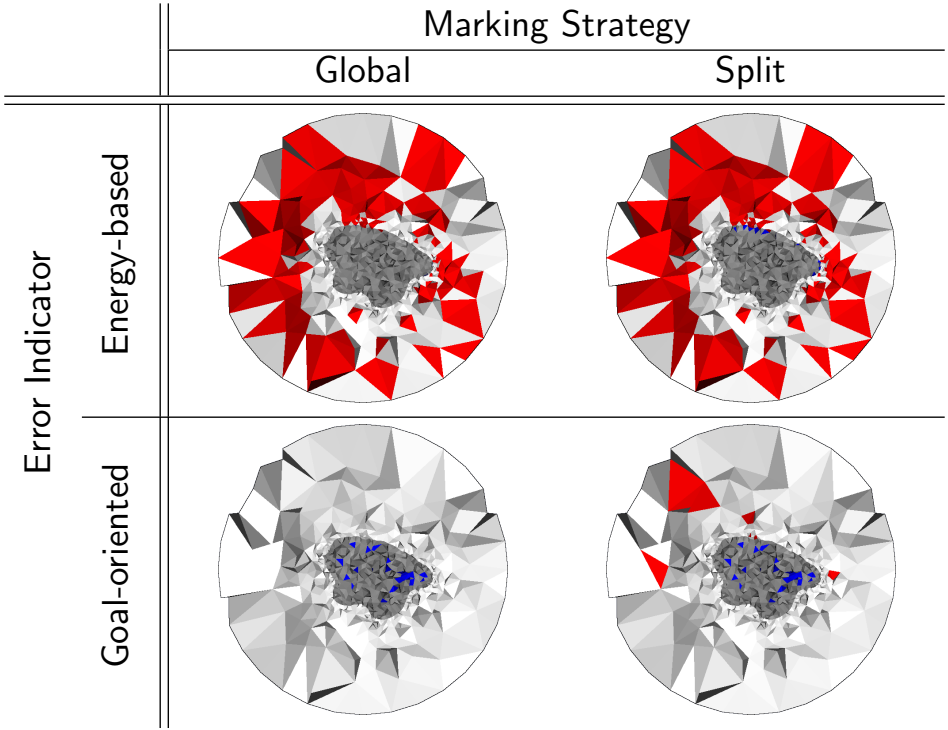
- "Exact" solution from uniform refinement
- Meshes from GAMer (Holst Group)
- Goal-oriented vs. energy-based refinement

Fasciculin-1



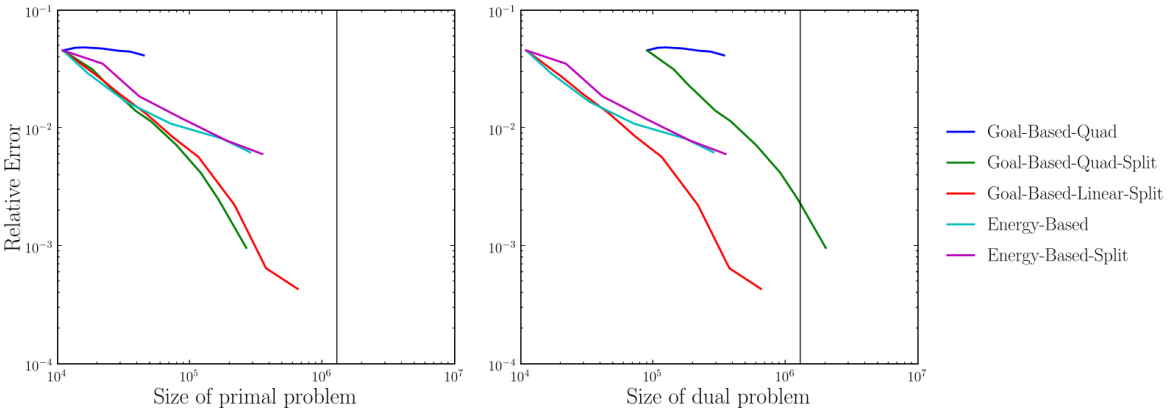
921 Atoms

Implicit Solvent Results



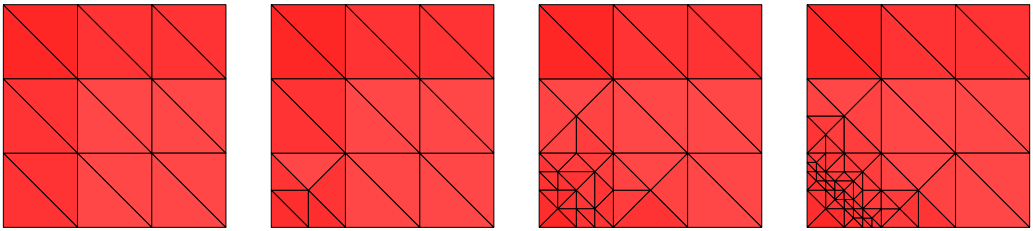
Implicit Solvent Results

Relative Error in Solvation Free Energy of Fasciculin-1



Take Home: Goal-oriented mesh refinement can achieve greater accuracy with fewer degrees of freedom

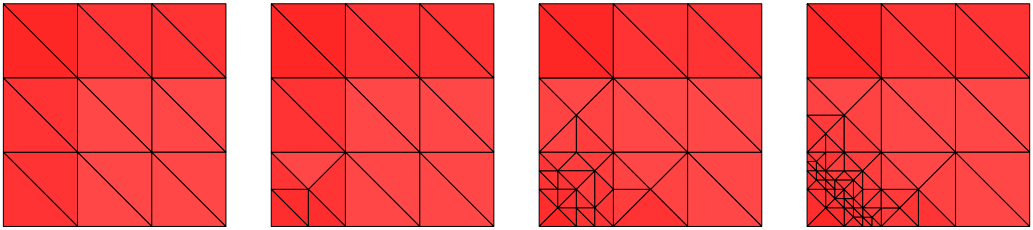
Multilevel Preconditioning: Multigrid



- Form a coarse problem: $A_{i-1} = P_i^t A_i P_i$
- Smooth: $A_i u_i \approx f_i, \quad r_i = f_i - A_i u_i$
- Restrict: $f_{i-1} = P_i^t r_i$
- Solve: $A_{i-1} u_{i-1} = f_{i-1}$
- Prolong: $u_i = u_i + P_i u_{i-1}$
- Smooth: $A_i u_i \approx f_i$

Electrostatics: Hierarchical Basis and BPX

B. Aksoylu, S. Bond, M. Holst, SIAM J. Sci. Comput. (2003)



- Introduce a change of basis
- Only smooth on the "new" mesh points
- Hierarchical Basis (Bank, Dupont, Yserentant)
  - Recursively defined locally supported basis functions
- BPX (Bramble, Pasciak, Xu)
  - Equivalent to smoothing on the "one-ring"

Outline

Final Thoughts

1. Poisson-Boltzmann equation models electrostatic effects of implicit solvent
2. Can develop error indicators using weak residual
3. Goal-oriented refinement requires the solution of dual problem
4. Solvation free energy accurately calculated using goal-oriented refinement
5. Appropriate marking strategy must be used
6. Multilevel preconditioning challenging with adaptively refined meshes

References

S.D. Bond and B.J. Leimkuhler, 'Molecular Dynamics and the Accuracy of Numerically Computed Averages,' *Acta Numerica*, 16 (2007) 1.

B. Aksoylu, S.D. Bond, E.C. Cyr, and M. Holst, Adaptive Solution of the PBE using Goal-Oriented Error Indicators. Preprint (2009)

S.D. Bond, J.H. Chaudhry, E.C. Cyr and L.N. Olson, 'A First Order Least Squares Finite Element Method for the PBE,' *J. Comput. Chem.*, submitted (2009).

Y. Cheng, J.K. Suen, Z. Radić, S.D. Bond, M.J. Holst and J.A. McCammon, 'Continuum Simulations of ACh Diffusion with Reaction-determined Boundaries in Neuromuscular Junction Models,' *Biophys. Chem.*, 127 (2007) 129.

Y. Cheng, J.K. Suen, D. Zhang, S.D. Bond, Y. Zhang, Y. Song, N. Baker, C.L. Bajaj, M. Holst and J.A. McCammon, 'Finite Element Analysis of the Time-Dependent Smoluchowski Equation for ACh Reaction Rate Calculations,' *Biophys. J.*, 92 (2007) 3397.

References

B. Lu, Y.C. Zhou, G.A. Huber, S.D. Bond, M.J. Holst and J.A. McCammon, 'Electrodiffusion: A continuum modeling framework for biomolecular systems with realistic spatiotemporal resolution' *J. Chem. Phys.*, 127 (2007) 135102.

K. Tai, S.D. Bond, H.R. MacMillan, N.A. Baker, M.J. Holst and J.A. McCammon, 'Finite Element simulations of ACh diffusion in Neuromuscular Junctions,' *Biophys. J.*, 84 (2003) 2234.

B. Aksoylu, S. Bond and M. Holst, 'Local refinement and multilevel preconditioning III: Implementation and Numerical Experiments,' *SIAM J. Sci. Comput.*, 25 (2003) 478.

S.D. Bond, B.J. Leimkuhler and B.B. Laird, 'The Nosé-Poincaré Method for Constant Temperature MD,' *J. Comput. Phys.*, 151 (1999) 114.