

An Efficient Monte Carlo Approach for Solving Linear Problems in Biomolecular Electrostatics

Charles Fleming¹, Michael Mascagni^{1,2}, and Nikolai Simonov^{2,3}

¹ Department of Computer Science and

² School of Computational Science,

Florida State University, Tallahassee, FL, 32306-4530, USA

³ ICM&MG, Lavrentjeva 6, Novosibirsk 630090, Russia

Abstract. A linear (elliptic) problem in molecular electrostatics is considered. To solve it, we propose an efficient Monte Carlo algorithm. The method utilizes parallel computing of point potential values. It is based on the walk-in-subdomains technique, walk-on-spheres algorithm, and an exact treatment of boundary conditions.

1 Introduction

Recently [1, 2], we proposed an advanced Monte Carlo technique for calculating different properties of large organic molecules. To use this approach, the problems under consideration have to be treated in the framework of continuous media models. In particular, to investigate electrostatic properties of molecules, we adopted one of the most popular and commonly used models, namely, the implicit solvent model. This means that solvent and ions dissolved within are treated as a continuous medium, whose properties are characterized by dielectric permittivity, ϵ_e , whereas the molecule under investigation is described explicitly. The solute (molecule) is thought of as a cavity with dielectric constant, ϵ_i , which is much less than that of the exterior environment.

Mathematically, this classical electrostatic approach leads to a boundary value problem for the Poisson equation satisfied by the electrostatic potential, $u(x)$:

$$-\nabla\epsilon\nabla u(x) = \rho(x), \quad x \in \mathbb{R}^3, \quad (1)$$

where ϵ is the position-dependent permittivity and $\rho(x)$ is the charge distribution [3]. Usually, the molecule under consideration is described geometrically as a union of spherical atoms with partial point charges at the sphere centers. This means that $\rho(x) = \sum_{m=1}^M q_m \delta(x - x^{(m)})$, where $q^{(m)}$ are the charges and $x^{(m)}$ are the atomic centers and hence the charge locations.

The charge distribution in the solvent is determined by the dissolved mobile ions. In the framework of a continuous medium approach, their positions in physical space are described by the Boltzmann statistical distribution; this leads to the non-linear Poisson-Boltzmann equation for $u(x)$ in the solvent [3].

For small potential values, this equation may be linearized, thus leading to the relation

$$\Delta u(x) - \kappa^2 u(x) = 0, \quad x \in G_1, \quad (2)$$

where κ^2 is a positive constant called the Debye-Hückel screening parameter.

All the difficulties that one encounters in solving such boundary-value problems arise from the geometrical features of the model of the molecule. The geometry comes through the boundary conditions that have to be satisfied by the limiting values of functions and their normal derivatives on different sides of the boundary, Γ :

$$u_i(y) = u_e(y), \quad \epsilon_i \frac{\partial u_i}{\partial n(y)} = \epsilon_e \frac{\partial u_e}{\partial n(y)}, \quad y \in \Gamma. \quad (3)$$

Here, Γ is either the surface of the molecule or the boundary of the so-called ion-exclusion layer. One of the possible variants is when these surfaces coincide.

One of the distinctive and attractive features of the Monte Carlo approach is that there is no need to perform cumbersome and otherwise labor-consuming preparatory approximations of the molecular surfaces and boundary conditions. In fact, a primitive analytic description based on a list of atomic centers and radii is absolutely adequate for type of random-walk simulations we carry out.

Another favorable aspect of the Monte Carlo algorithms we employ is their natural, immanent parallelizability. For linear electrostatic problems, the desired computational entity can often be represented as a linear functional of point potential values. This makes possible further distribution of the calculations and therefore computing these values in parallel.

2 Energy: Linear Case

We consider the problem of calculating the electrostatic free energy of a large biomolecule. In the linear case, this energy is equal to

$$E = \frac{1}{2} \sum_{m=1}^M u^{(m)} q^{(m)}, \quad (4)$$

where $u^{(m)}$ is the regular part of the electrostatic potential at the center of the m th atom. To estimate $u^{(m)}$, $m = 1, \dots, M$ we make use of the walk-on-spheres algorithm. The specific form of the molecule's geometrical model makes it possible to simulate the required Markov chains in such a way that the chain converges to the boundary geometrically, and the last point lies exactly on Γ .

Let $G_i = \bigcup_{m=1}^M B(x^{(m)}, r^{(m)})$ be the inside of the molecule. Here, $B(x, r)$ denotes the ball centered at the point x , and r is its radius (see Figure 1). For a given distribution of point charges in G_i , the electrostatic potential can be

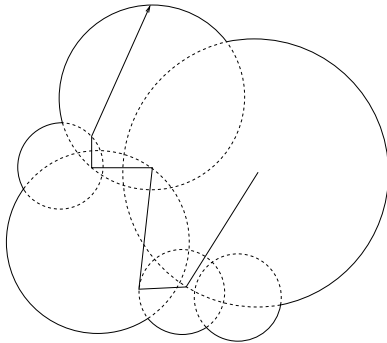


Fig. 1. An example of a computational domain in the energy calculation. The domain is a union of atomic spheres, and the line follows a walk-on-spheres/walk-in-subdomains trajectory that begins at the center of the largest sphere and ends as it leaves the interior of the computational domain

represented as $u_i(x) = u^0(x) + g(x)$, where $g(x) = \sum_{m=1}^M \frac{q^{(m)}}{4\pi\epsilon_i} \frac{1}{|x - x^{(m)}|}$. From

this it follows that the regular part of the electrostatic potential, u^0 , satisfies the Laplace equation in G_i . To find $u^{(m)} = u^0(x^{(m)})$, we construct a Monte Carlo estimate based on combination of the walk-on-spheres and walk-in-subdomains algorithms [4]. Let $\xi[u^0](x_0)$ be a Monte Carlo estimate for $u^0(x_0)$. Using the probabilistic representation, we set $\xi[u^0](x_0) = u^0(x^*)$, where x^* is the exit point of Brownian motion from G . The most efficient way to simulate these exit points is to use the natural representation of G as a union of intersecting spherical subdomains. For every ball, the exit point can be sampled directly. By the Poisson integral formula for a function, u , that satisfies the Laplace equation, at every point $x \in B(x^{(m)}, r^{(m)})$ we have $u(x) = \int_{S(x^{(m)}, r^{(m)})} p_p(x \rightarrow y) u(y) d\sigma(y)$, where $p_p(x \rightarrow y) = \frac{1}{4\pi r^{(m)}} \frac{(r^{(m)})^2 - |x - x^{(m)}|^2}{|x - y|^3}$ is the Poisson kernel. The Markov chain $\{x_i, i = 0, 1, \dots, N\}$ is defined recursively. Given $x_i \in B(x_i^{(m)}, r_i^{(m)})$, the next point is chosen randomly on $S(x_i^{(m)}, r_i^{(m)})$ in accordance with density $p_p(x_i \rightarrow x_{i+1})$. By definition, x_{i+1} either hits the boundary or lies in some other ball. With probability one, the path of this Markov chain terminates at the boundary after a finite number of steps, N . With some natural assumptions about G , which are always valid for molecular geometries, $\mathbb{E}N < \infty$. We can then set $x^* = x_N$.

3 Boundary Conditions

Next, we have to construct estimates for boundary values of the potential, u^0 . By definition $\xi[u^0](x_N) = \xi[\psi](x_N) - g(x_N)$, where ψ is the unknown boundary

value of u . One approach for constructing a Monte Carlo estimate for ψ is to randomize a finite-difference approximation of the boundary condition with step-size h [1]. Hence, we have $\psi(x_N) = \mathbb{E}(u(x_{N+1})|x_N) + O(h^2)$, where $x_{N+1} = x_N - hn$ with probability $p_i = \frac{\epsilon_i}{\epsilon_i + \epsilon_e}$, $x_{N+1} = x_N + hn$ with the complementary probability, p_e , and n is the external normal vector. This introduces an $O(h)$ bias into resulting Monte Carlo estimate.

Here, we propose another approach that speeds up the computations and eliminates the bias. This approach is new and is specifically adapted to the boundary flux-type conditions that include normal derivatives. Consider an auxiliary sphere $S(x_N, a)$, which does not contain point charges. Then, application of Green's formula to functions u_i, u_e and $\Phi_\kappa(x) = -\frac{1}{4\pi} \frac{\sinh(\kappa(a-r))}{r \sinh(\kappa a)}$, where $r = |x - x_N|$, leads to the following mean-value relation:

$$\begin{aligned}
 u(x_N) &= \frac{\epsilon_e}{\epsilon_e + \epsilon_i} \int_{S_e(x_N, a)} \frac{1}{2\pi a^2} \frac{\kappa a}{\sinh(\kappa a)} u_e \\
 &+ \frac{\epsilon_i}{\epsilon_e + \epsilon_i} \int_{S_i(x_N, a)} \frac{1}{2\pi a^2} \frac{\kappa a}{\sinh(\kappa a)} u_i \\
 &- \frac{(\epsilon_e - \epsilon_i)}{\epsilon_e + \epsilon_i} \int_{\Gamma \cap B(x_N, a) \setminus \{x_N\}} 2 \frac{\partial \Phi_0}{\partial n} Q_{\kappa, a} u \\
 &+ \frac{\epsilon_i}{\epsilon_e + \epsilon_i} \int_{B_i(x_N, a)} [-2\kappa^2 \Phi_\kappa] u_i,
 \end{aligned} \tag{5}$$

where $Q_{\kappa, a}(r) = \frac{\sinh(\kappa(a-r)) + \kappa r \cosh(\kappa(a-r))}{\sinh(\kappa a)} < 1$, and Γ is the surface

of the molecule.

Randomization of this formula provides a procedure to extend the Markov chain $\{x_i\}$ after it hits the boundary without introducing any bias. The algorithm works as follows. With probability p_e the next point is chosen in the exterior of the molecule, \overline{G}_e , and with the complementary probability, p_i , $x_{N+1} \in \overline{G}_i$. In the exterior, the next point of the Markov chain is chosen isotropically on the boundary of $B(x_N, a) \cap G_e$. The random walker in this case has non-zero probability of terminating. On $S(x_N, a)$ the termination probability is equal to $1 - \frac{\kappa a}{\sinh(\kappa a)}$. Note that $B(x_N, a) \cap G_e$ is not convex. This means that there can be two intersections with its boundary. In this case we can choose randomly one of them. In the interior, we choose an isotropic direction first. Then, with the corresponding probabilities, the next point is sampled on the chosen straight line either on the boundary of $B(x_N, a) \cap G_i$, or inside the domain. From (5) it follows that $\psi(x_N) = \mathbb{E}u(x_{N+1})$.

4 Random Walks Inside and Outside

The further construction is common to both methods of treating these boundary conditions.

For internal points, we construct a Markov chain of the walk-in-subdomains starting at x_{N+1} , and set $u(x_{N+1}) = \mathbb{E}(u^0(x_2^*) + g(x_{N+1}) | x_{N+1})$ where $x_2^* \in \Gamma$ is the next exit point.

For $u(x_{N+1}) \in G_e$, we can use the walk-on-spheres algorithm. A random walker of this Markov chain survives at every step with probability $\frac{\kappa d_i}{\sinh(\kappa d_i)}$, where d_i is the distance to the boundary. With probability one this chain terminates. This happens either inside G_e or when the distance d_i from the point to the boundary becomes less than a prescribed number, ε . In the second case we return to estimating ψ on the boundary. Since the probability of terminating is bounded away from zero, the mean number of steps in the Markov chain is finite.

The same approach works also for the case when the ion-exclusion layer is introduced into the molecule model. This means that the whole space is divided into three domains: the internal part of the molecule, G_i , the external, G_e , and the layer between them, G_0 . In G_0 we have $\epsilon = \epsilon_e$ and $\kappa = 0$. On ∂G_i the mean-value relationship (5) then reads:

$$\begin{aligned} u(x_N) = & \frac{\epsilon_e}{\epsilon_e + \epsilon_i} \int_{S_e(x_N, a)} \frac{1}{2\pi a^2} u_e \\ & + \frac{\epsilon_i}{\epsilon_e + \epsilon_i} \int_{S_i(x_N, a)} \frac{1}{2\pi a^2} u_i \\ & - \frac{\epsilon_e - \epsilon_i}{\epsilon_e + \epsilon_i} \int_{\Gamma \cap B(x_N, a) \setminus \{x_N\}} 2 \frac{\partial \Phi_0}{\partial n} u, \end{aligned} \quad (6)$$

where $\Phi_0(x) = -\frac{1}{4\pi} \left(\frac{1}{|x - x_N|} - \frac{1}{a} \right)$. On ∂G_e the dielectric constants are equal on both sides. This means that the third term in (5) equals zero.

5 Experiments, Conclusions, and Future Work

To test the proposed algorithms we applied our Monte Carlo estimates to a simple model problem with a known analytic solution. For a spherical (one-atom) molecule, the exact value of E scaled by $q^2/(4\pi\epsilon_i R)$ equals $\epsilon_i/(\epsilon_e(1 + \kappa R)) - 1$. For the values $\epsilon_i = 4.0$, $\epsilon_e = 78.5$, $\kappa = 0$ one obtains the exact solution of -0.9490 . Our calculation based on the walk-on-spheres method provided the result -0.9489 with a 0.1% statistical error in 0.7 seconds for $a = 0.1$ and in 1.4 seconds for $a = 0.03$. The algorithm based on the finite-difference approximation of the boundary condition with $h = 0.001$ provided the result with the same accuracy in 5.6 seconds. All calculations were carried out on an ordinary desktop

computer with a 1.3 GHz P4 processor running Windows 2000, which is already a fairly slow computer.

Thus we see roughly a speed increase of one order of magnitude when progressing from the finite-difference approach to the boundary condition to this new method. In addition, this comes with the elimination of the $O(h)$ bias of the finite-difference technique. While we carried out these very simple computations on an ordinary PC, it is important to note that the specific form of the functional (4) we computed makes it possible to simulate in parallel M independent random walks for estimating every $u^{(m)}$ very much more rapidly in a parallel setting.

We believe these techniques will be very important in the eventual creation of a suite of Monte Carlo approaches to problems and computations in biomolecular electrostatics. Thus, we plan to continue this line of research towards this long-term goal. In the short term we plan to investigate other molecular surfaces and their implications for these methods, the direct solution of the nonlinear Poisson-Boltzmann equation via Monte Carlo, and the calculation of forces and other types of functionals of the potential and its partial derivatives.

References

1. Mascagni, M. and Simonov, N.A.: Monte Carlo method for calculating the electrostatic energy of a molecule. In: Lecture Notes in Computer Science, Vol. 2657. Springer-Verlag, Springer-Verlag, Berlin Heidelberg New York (2003) 63–74.
2. Mascagni, M. and Simonov, N.A.: Monte Carlo methods for calculating some physical properties of large molecules. *SIAM Journal on Scientific Computing*. **26** (2004) 339–357.
3. Davis, M.E. and McCammon, J.A.: Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **90** (1990) 509–521
4. Simonov, N.A.: A random walk algorithm for the solution of boundary value problems with partition into subdomains. in *Metody i algoritmy statisticheskogo modelirovaniya*, Akad. Nauk SSSR Sibirsk. Otdel., Vychisl. Tsentr, Novosibirsk (1983) 48–58.