

Random walk algorithms for estimating electrostatic properties of large molecules

Nikolai A. Simonov ^{*}, Michael Mascagni [†]

Abstract

The problem of computing the electrostatic energy of a large molecule is considered. It is reduced to solving the Poisson equation inside the domain and the linear Poisson-Boltzmann equation in the exterior, coupled by boundary conditions. Monte Carlo estimates for the potential point values, their derivatives, and the energy are constructed. These estimates are based on the walk on spheres and walk on boundary algorithms. Results of some illustrative calculations are presented.

1 Introduction

Electrostatics plays a fundamental role in intermolecular interactions and to a large extent determines molecular properties [1, 10]. There are different approaches to the description of the electrostatic field on the molecular level. One of the possible and widely used models is a continuum model. For a given charge distribution $\rho(x)$, the electrostatic potential is determined as a solution to Poisson's equation

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

where ϵ is a position-dependent permittivity. In bio-molecular applications, the geometry of a problem is taken into account by thinking of a molecule as a cavity with point charges and constant ϵ inside. The exterior is considered as a dielectric medium with different permittivity and some charge distribution. Clearly, certain boundary conditions on the surface of a molecule have to be added.

^{*}Institute of Computational Mathematics and Mathematical Geophysics SB RAS and School of Computational Science and Information Technology, Florida State University, Tallahassee, FL, 32306, USA

[†]Department of Computer Science and School of Computational Science and Information Technology, Florida State University, Tallahassee, FL, 32306, USA

The common approach to solving these kinds of problems is the finite-difference technique; boundary element and finite element methods are successfully used as well.

An alternative way of treating computationally boundary value problems is to apply Monte Carlo methods. For elliptic equations, the most efficient and commonly used are the walk on spheres (WOS) method [9] and walk on boundary algorithm [11]. It is worth noting, however, that these methods are well established for the equations with constant coefficients and classical boundary conditions. In molecular electrostatics problems, we have to deal with coupling different equations through continuity boundary conditions.

2 Mathematical Statement of the Problem

In calculating the internal electrostatic energy of a molecule, we accept the model which is commonly used in biomolecular electrostatics computations [1]. The molecule in question is considered as a union of a large number of intersecting balls (atoms) that constitutes a compact set $G \in \mathbb{R}^3$:

$$G = \bigcup_{m=1}^M B(x_m, r_m) .$$

Every spherical atom has its electrical charge, q_m , which is positioned at its center, x_m , and r_m is the radius of this atom (ball). Hence, the electrostatic potential, $u(x)$, satisfies Poisson's equation (1) inside G for the particular

charge density $\rho(x) = \sum_{m=1}^M q_m \delta(x - x_m)$. Here, the dielectric permittivity, $\epsilon = \epsilon_i$, is constant.

Assume first that the molecule is grounded. Physically, this means that G is surrounded by an ideally conducting material, and the problem of computing the potential reduces to the interior Dirichlet boundary value problem for Poisson's equation (1): $u(y) = 0$.

In the general case, the molecule can be thought of as being surrounded by some dielectric (e.g. water). The classical approach is to treat the exterior medium as continuous with some constant permittivity, ϵ_e . The distribution of dissolved ions determines the charge density outside G . From statistical mechanics considerations, these ions should be distributed in accordance with the Boltzmann law. For small potential values, the resulting equation may be linearized, thus leading to the linear Poisson-Boltzmann equation:

$$\Delta u(x) - \kappa^2 u(x) = 0 , \quad x \in G_1 \equiv \mathbb{R}^3 \setminus \overline{G} , \quad (2)$$

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

Equations (1) and (2) must be coupled by the continuity conditions on the boundary of the molecule:

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

For convenience, we denote u_i as the solution to (1) in the interior of G , and u_e as the solution of (2) in the exterior of the molecule. We also assume that $u_e(x) \rightarrow 0$ as $|x|$ goes to infinity.

In the linear case, the electrostatic free energy of the molecule is given by [1, 10]

$$E = \frac{1}{2} \sum_{m=1}^M u_m q_m ,$$

where u_m is the non-singular part of the electrostatic potential at the center of the m -th atom. This means that in calculation of E we exclude the infinite self-energy of point charges, and take $u_m = u^{(0)}(x_m)$. Here, $u^{(0)}$ comes from the available explicit decomposition of the potential inside G :

$$u(x) = u^{(0)}(x) + g(x) , \quad (4)$$

where $g(x) = \sum_{m=1}^M \frac{q_m}{4\pi\epsilon_i} \frac{1}{|x - x_m|}$.

From (4) and (1) it follows that $\Delta u^{(0)}(x) = 0$ in G , and the boundary values of $u^{(0)}$ are equal to $u - g$.

Summary. There are two main problems that we have to address here. The first is the discontinuity of ϵ that gives rise to the boundary condition (3). The second is the discontinuity of the coefficient κ . Here, we assume that both these discontinuities occur on the boundary of a molecule. Note that in other commonly used models, the discontinuity surfaces for ϵ and κ are different.

3 Monte Carlo Algorithm: Walk in Subdomains

Let $\xi[E]$ be a Monte Carlo estimate for E . We represent it as a weighted sum of estimates for point values of the potential:

$$\xi[E] = \frac{1}{2} \sum_{m=1}^M \xi[u_m] q_m .$$

Using the probabilistic representation [6, 3], we set $\xi[u_m] = u^{(0)}(x_m^*)$, where x_m^* are exit points 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 [12]. For every ball, the exit point can be sampled directly. By Poisson's formula for a function, u , that satisfies the Laplace equation, at every point $x \in B(x_c, r)$ we have

$$u(x) = \int_{S(x_c, r)} p_p(x \rightarrow y) u(y) d\sigma(y), \text{ where } p_p(x \rightarrow y) = \frac{1}{4\pi r} \frac{r^2 - |x - x_c|^2}{|x - y|^3}$$

is the Poisson kernel. For every m , we set $x_m^{(0)} = x_m$ and define the Markov chain $\{x_m^{(i)}, i = 0, 1, \dots, N_m\}$ recursively. Given $x_m^{(i)}$, the next point is chosen randomly on $S(x_m^{(i)}, r_m^{(i)})$ with density $p_p(x_m^{(i)} \rightarrow x_m^{(i+1)})$. By definition, $x_m^{(i+1)}$ 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_m . We can then set $x_m^* = x_m^{(N_m)}$. For a grounded mole-

cule, $u = 0$ on the boundary, and therefore $\xi[E] = -\frac{1}{2} \sum_{m=1}^M g(x_m^{(N_m)}) q_m$.

To evaluate the energy in the general case, we have to construct estimates for boundary values of the potential, $u^{(0)}$. By definition

$$\xi[u^{(0)}](y) = \xi[\psi](y) - g(y), \quad y \in \partial G, \quad (5)$$

where $\psi(y)$ is the unknown boundary value of $u(y)$. 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 [8]. This introduces an $O(h)$ bias into the solution.

Here, we propose another approach [7] that speeds up the computations and to a large extent eliminates the bias. The derivation of this new procedure begins by assuming $\kappa = 0$. Also assume that the boundary is flat, and consider a sphere $S(y, a)$ for some a with $y \in \partial G$. Then using separation of variables makes it possible to prove that despite the discontinuity of the normal derivative, the mean value property holds:

$$\psi(y) = p_e \int_{S_e(y, a)} u_e / |S_e(y, a)| + p_i \int_{S_i(y, a)} u_i / |S_i(y, a)|. \quad (6)$$

Here $p_e = \epsilon_e / (\epsilon_e + \epsilon_i)$, $p_i = 1 - p_e$, $S_e \subset G_1$ and $S_i \subset G$. $|S_e(y, a)|$ denotes the area of the part of the sphere extending into the exterior. In this case it equals $2\pi a^2$. For a spherical boundary (6) is not exact, but it still holds to high accuracy if we take $p_e = \epsilon_e |S_e(y, a)| / (\epsilon_e |S_e(y, a)| + \epsilon_i |S_i(y, a)|)$ (simulation-tabulation approach can also be used in this case [5]).

The algorithm works as follows. With probability p_e the next point, y_1 , is chosen isotropically on S_e and $u(y_1) = u_e(y_1)$; with the complementary

probability y_1 is chosen isotropically on S_i and $u(y_1) = u_i(y_1)$. From (6) it follows that $\psi(y) = \mathbb{E}u(y_1) + \varsigma$, where small bias ς depends on a/r_m . For internal points, we construct Markov chain of the walk in subdomains starting at y_1 and set $u(y_1) = \mathbb{E}(u^{(0)}(y_2) + g(y_1)|y_1)$ where $y_2 \in \partial G$ is the exit point.

For $y_1 \in G_1$, we can use the walk on spheres algorithm [2]. We set $y_{1,0} = y_1$ as the starting point. The WOS Markov chain terminates either when the distance d_i from the point $y_{1,i}$ to the boundary becomes less than a prescribed number, ε , or according to the Poisson kernel in the exterior of a sphere the point is set to infinity ($\psi = 0$). In the first case, we return to estimating ψ on the boundary.

4 Monte Carlo Algorithm: Walk on the Boundary

In this section we consider another Monte Carlo algorithm that can be used for solving the boundary value problem (1), (3). Assume that ∂G is smooth enough, to represent the non-singular part of the solution in the form of a single-layer potential [4]: $u^{(0)}(x) = \int_{\partial G} \frac{1}{2\pi} \frac{1}{|x-y|} \mu(y) d\sigma(y)$. Taking into account the boundary conditions and discontinuity properties of the single-layer potential's normal derivative [4], we arrive at the integral equation for the unknown density, μ :

$$\mu = -\lambda_0 \mathcal{K} \mu + f, \quad (7)$$

which is valid almost everywhere on ∂G . Here, $\lambda_0 = \frac{\epsilon_e - \epsilon_i}{\epsilon_e + \epsilon_i}$, and the kernel of the integral operator \mathcal{K} is $\frac{1}{2\pi} \frac{\cos \phi_{yy'}}{|y-y'|^2}$, where $\phi_{yy'}$ is the angle between the external normal vector $n(y)$, and the vector $y - y'$; $f = \lambda_0 \frac{\partial g}{\partial n(y)}$ and can be computed analytically. Since $\lambda_0 < 1$, the Neumann series for (7) converges (see, e.g. [4]), and it is possible to calculate the solution as

$$u^{(0)}(x) = \sum_{i=0}^{\infty} (h_x, (-\lambda_0 \mathcal{K})^i f), \quad (8)$$

where $h_x(y) = \frac{1}{2\pi} \frac{1}{|x-y|}$. Usually, however, $\epsilon_e \gg \epsilon_i$ and, hence, $|\lambda_1 - 1| = \frac{2\epsilon_i}{\epsilon_e + \epsilon_i} \ll 1$. Here, $\lambda_1 = -1/\lambda_0$ is the largest eigenvalue of the operator $(-\lambda_0 \mathcal{K})^{-1}$. This means that convergence in (8) is rather slow.

To speed up the convergence in (8), we apply the method of spectral parameter substitution. In this particular case, it is possible to use the function $\lambda = \frac{2|\lambda_1|\eta}{1-\eta}$, and hence

$$u^{(0)}(x) = \sum_{i=0}^n l_i^{(n)} (-\lambda_0)^i (h_x, \mathcal{K}^i f) + O(q^{n+1}), \quad (9)$$

where $q = \frac{1}{1+2|\lambda_1|} < \frac{1}{3}$, and $l_i^{(n)} = \sum_{j=i}^n \binom{i-1}{j-1} (2|\lambda_1|)^i q^j$. Given a desired computational accuracy, we can calculate the number of terms needed in (9). Thus, the problem reduces to computing a finite number of multidimensional integrals $I_i(x) = (h_x, \mathcal{K}^i f)$.

For a convex G , we can use the isotropic random walk on boundary Markov chain $Y = \{y_0, y_1, \dots\}$ [11]. Therefore $I_i(x) = \mathbb{E}Q_i h_x(y_i)$, where random weights are $Q_i = \frac{f(y_0)}{p_0(y_0)}$, $i = 1, 2, \dots, n$. Hence, q^{n+1} -biased estimator for

$$u^{(0)} \text{ is } \theta = \sum_{i=0}^n l_i^{(n)} (-\lambda_0)^i \frac{f(y_0)}{p_0(y_0)} h_x(y_i).$$

5 Results of Computational Experiments

To test the proposed algorithms we applied our Monte Carlo estimates to a simple model problem with known analytical solution. All calculations were carried out on an ordinary desktop computer: a Dell PC with 1.3 GHz P4 processor running Windows 2000.

For the simplest case, a spherical (one-atom) molecule, the analytical solution scaled by $\frac{q^2}{4\pi\epsilon_i R}$ is $E = \frac{\epsilon_i}{\epsilon_e(1+\kappa R)} - 1$. For the values $\epsilon_i = 4.0$, $\epsilon_e = 78.5$, $\kappa = 0$ one obtains the exact solution, -0.9490 . Our calculation based on the walk on spheres provided the result -0.9494 in 3 seconds. For $a = 0.2$ and number of samples $N = 4000$, the statistical error σ (two standard errors) is equal to 9.6×10^{-4} (i.e. 0.1%). Note that there is no bias, since mean value formula (6) is exact for this particular case. The algorithm based on the finite-difference approximation of the boundary condition provided the result -0.9486 in 29 seconds ($h = 0.001$, $N = 15000$, $\sigma = 9.5 \times 10^{-4}$ and bias is $O(h)$). For a spherical molecule, the walk on the boundary algorithm gives the exact solution biased by $O(q^{n+1})$.

References

- [1] Davis M.E. and McCammon J.A. Electrostatics in biomolecular structure and dynamics // *Chem. Rev.* – 1990. – Vol. 90. – P. 509-521.
- [2] Elepov B.S., Kronberg A.A., Mikhailov G.A. and Sabelfeld K.K. Solution of boundary value problems by the Monte Carlo method. – Novosibirsk: Nauka, 1980 (in Russian).
- [3] Freidlin M. Functional integration and partial differential equations. – Princeton: Princeton University Press, 1985.
- [4] Günter, N.M. Potential theory and its application to the basic problems of mathematical physics. – Moscow: Gostehizdat, 1953 (in Russian).
- [5] Hwang C.-O., Given J.A. and Mascagni M. The simulation-tabulation method for classical diffusion Monte Carlo // *J. Comput. Phys.* – 2001. – Vol. 174. – P. 925-946.
- [6] Kac M.H. On some connections between probability theory and differential and integral equations // In: *Proc. 2nd Berkeley Symp. on Math. Statist. and Probability.* – Berkeley: Univ. of California Press, 1951. P. 189-215.
- [7] In Chan Kim and Torquato S. Determination of the effective conductivity of heterogeneous media by Brownian motion simulation // *J. Appl. Phys.* – 1990. – Vol. 68. – P. 3882-3903.
- [8] 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. – Heidelberg: Springer-Verlag, 2003. P. 63-74.
- [9] Muller M.E. Some continuous Monte Carlo methods for the Dirichlet problem // *Ann. Math. Statist.* – 1956. – Vol. 27. – P. 569-589.
- [10] Naray-Szabo, G. and Warshel, A. (eds.) *Computational Approaches to Biochemical Reactivity*. Series: *Understanding Chemical Reactivity*, Vol. 19. – New York: Kluwer Academic Publishers, 2002.
- [11] Sabelfeld K.K. and Simonov N.A. *Random Walks on Boundary for solving PDEs.* – Utrecht: VSP, 1994.
- [12] Simonov, N.A. A random walk algorithm for the solution of boundary value problems with partition into subdomains // In: *Methods and algorithms for statistical modelling.* – Novosibirsk: Akad. Nauk SSSR Sibirsk. Otdel., Vychisl. Tsentr, 1983. – P. 48-58 (in Russian).