Lietuvos matematikos rinkinys. LMD darbai
51 tomas, 2010, 250–255

ISSN 0132-2818
www.mii.lt/LMR/

The influence of bounding surface on the precision of the Cauchy–Crofton method

Justinas V. Daugmaudis[1], **Audrius Laurynėnas**[2], **Feliksas Ivanauskas**[1]

[1] *Vilnius University, Faculty of Mathematics and Informatics*
Naugarduko 24, LT-03225 Vilnius
[2] *Vilnius University, Faculty of Chemical*
Naugarduko 24, LT-03225 Vilnius
E-mail: jvd@binf.ku.dk; audrius.laurynenas@chf.stud.vu.lt; feliksas.ivanauskas@maf.vu.lt

Abstract. We derive and implement a method to compute isosurface area that demonstrates better numeric properties than previously described similar algorithms. The described stochastic area computation algorithm has been tested on geometric objects as well as on protein models. One of the advantages of the method is that for each line in a sample the number of surface intersection points can be counted in a parallel manner, independently of other lines in the sample, which maps well to the multi-core and multi-node architectures.

Keywords: Cauchy–Crofton method, bounding surfaces, molecular surface measurement, stochastic algorithm.

Introduction

In order to solve many practical problems that arise in fundamental sciences, it is often necessary to measure isosurface area. For instance, one of the most challenging questions in biochemistry is how to compute the ternary structure of proteins from the sequence of aminoacids. When this question is answered fully, many other problems concerning, for example, rational mutagenesis, drug design, protein-ligand interaction, the creation of biocatalysts and biotechnological methods will find their solution, too. By employing molecular mechanics and dynamics to model the ternary structure of proteins *ab initio*, we have to continually compute the area of protein surface. The modelling speed depends on the speed of the surface area computations. Also, whether the modelling result will match experimental data depends on the precision of those computations.

1 Stochastic measurement of the surface area

We can use the Cauchy–Crofton formula to measure the surface area, as in [2, 1]. Let's assume that Γ is a given surface, and we have to compute its area $S(\Gamma)$, for which the following formula is valid:

$$\int_{D\in L} \#(\Gamma \cap D)\, dL = \pi S(\Gamma). \tag{1}$$

Here S is the function of surface area, $\#(\Gamma \cap D)$ is a number of intersection points between a line $D \in L$ and the surface Γ. L is a set of random lines. You can find the detailed proof of the (1) formula in [2]. Li et al. used the Cauchy–Crofton formula to compute the surface area of CGS (Computer-generated scene) models by employing the Monte Carlo integration approximation. The method relates the surface area of a 3D body A with the number of intersection points between A's bounding surface B and the set of random lines in $\mathbb{R}^3$.

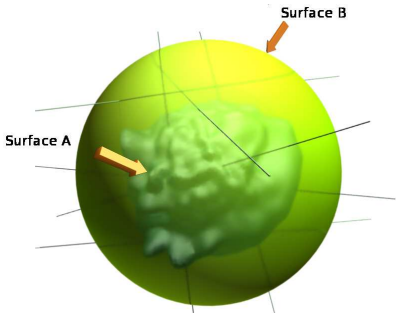


Fig. 1: The area of the surface A that is inscribed within the bounding surface B can be computed using the Cauchy–Crofton method. Random lines are generated within B.

Let's assume that we are examining a given surface A and we have to compute its area $S(A)$. B is a bounding surface for which the area is given by $S(B)$. The area $S(B)$ of the surface B is known. Let $\mathcal{L}$ be a set of random lines that intersect the surface B as shown in Fig. 1. The surface B will be intersected exactly $N = 2|\mathcal{L}|$ times; each line intersects the bounding surface 2 times – at the entrance and exit points. Suppose that $n_A = \sum_{D\in\mathcal{L}} \#(A \cap D)$ and $n_B = \sum_{D\in\mathcal{L}} \#(B \cap D)$ are, respectively, the total number of intersection points between A and B with lines in the set $\mathcal{L}$. Employing the integration approximation for the (1) formula, Li et al. state that $\frac{n_A}{N} \approx c\pi S(A)$ and $\frac{n_B}{N} \approx c\pi S(B)$. Here, c is a proportional constant. Now the unknown proportional constant c can be eliminated, and the area of the surface A is computed using the formula:

$$S(A) \approx \frac{n_A}{n_B} S(B). \tag{2}$$

Based on the (2) formula, the method used by Li et al. is, essentially, one of the Monte Carlo integration approximation methods.

2 Partitioning of surfaces

Let f be a random variable on the probability space (A, B, λ). We partition A into sets $\Omega_1, \ldots, \Omega_k \in B$, where $\lambda(\Omega_j) > 0$, $j = \overline{1,k}$. For all $j = \overline{1,k}$ we choose N_j independent random samples $\xi_1^{(j)}, \ldots, \xi_{N_j}^{(j)} \in \Omega_j$ with the probability measure $\mu_j = \lambda(\Omega_j)^{-1}\lambda$. Let $I(f) = \int_B f\, d\lambda$ be the expected value of the random variable. Then the following formulas are valid:

$$I(f) = \sum_{j=1}^{k} \int_{\Omega_j} f\, d\lambda = \sum_{j=1}^{k} \lambda(\Omega_j) \int_{\Omega_j} f\, d\mu_j. \tag{3}$$

From (3) we can derive that the Monte Carlo estimate for all partitions Ω_j is given by these formulas:

$$\lambda(\Omega_j)\int_{\Omega_j} f\,d\mu_j \approx \frac{\lambda(\Omega_j)}{N_j}\sum_{n=1}^{N_j} f\big(\xi_n^{(j)}\big), \quad j=\overline{1,k}.$$

From this obviously follows that

$$I(f) \approx \sum_{j=1}^{k}\frac{\lambda(\Omega_j)}{N_j}\sum_{n=1}^{N_j} f\big(\xi_n^{(j)}\big). \tag{4}$$

3 Generalized Cauchy–Crofton formula

From the (3) formula we derive that the partitioning of an arbitrary surface Γ into $k \geqslant 1$ parts, so that $\Gamma_1,\ldots,\Gamma_k \in \Gamma$, where $S(\Gamma_j) > 0$, $j=\overline{1,k}$, and $\sum_{n=1}^{k} S(\Gamma_n) = S(\Gamma)$, is given by

$$\sum_{n=1}^{k}\int_{D\in L_n} \#(\Gamma_n \cap D)\,dL_n = \pi S(\Gamma). \tag{5}$$

We can show that the (5) formula offers a better precision than (1) because the following inequality

$$\begin{aligned}&\int_{\Omega}\cdots\int_{\Omega}\left(\sum_{j=1}^{k}\frac{\lambda(\Omega_j)}{N_j}\sum_{n=1}^{N_j} f\big(\xi_n^{(j)}\big) - I(f)\right)^2 d\lambda(\xi_1)\ldots d\lambda(\xi_N),\\ &\quad\leqslant \int_{\Omega}\cdots\int_{\Omega}\left(\frac{1}{N}\sum_{n=1}^{N} f(\xi_n) - I(f)\right)^2 d\lambda(\xi_1)\ldots d\lambda(\xi_N)\end{aligned}$$

is satisfied, when $N=\sum_{j=1}^{k} N_j$ and $N_j = \lambda(\Omega_j)N$.

We can interpret this statement in two ways: either we get a better precision or, using smaller random line samples, the same precision for shorter computation times.

From the condition $N_j = \lambda(\Omega_j)N$ we state that

$$N_j = \frac{S(\Gamma_j)}{S(\Gamma)}N, \quad j=\overline{1,k}. \tag{6}$$

Obviously, the requirement $N=\sum_{j=1}^{k} N_j$ is satisfied because

$$\sum_{j=1}^{k}\frac{S(\Gamma_j)}{S(\Gamma)}N = \frac{N}{S(\Gamma)}\sum_{j=1}^{k} S(\Gamma_j) = N.$$

The (5) formula is a generalization of the Cauchy–Crofton method, and the (1) formula is the specific case of (5) with $k=1$.

4 Condition to terminate the computations

Let $\mathcal{L}_1 \dots \mathcal{L}_k$ be random samples of lines that intersect, respectively, surfaces $\Gamma_1 \dots \Gamma_k$, $\#(\Gamma_n \cap D)$ is a number of intersection points between a line $D \in \mathcal{L}_n$ and the surface Γ_n. From the (5) formula we derive that

$$\lim_{|\bigcup_{i=1}^{k} \mathcal{L}_i| \to \infty} \left(\pi S(\Gamma) - \sum_{n=1}^{k} \sum_{D \in \mathcal{L}_n}^{|\mathcal{L}_n|} \#(\Gamma_n \cap D) \right) = 0. \tag{7}$$

Let's assume that $1 \ll N < \infty$ iterations have been carried out, the necessary condition is that $\mathcal{L}_i^{(N)} \supset \mathcal{L}_i^{(N-1)} \supset \cdots \supset \mathcal{L}_i^{(1)}$, $i = \overline{1,k}$. Let $1 < m < N$ be the number of the last sequential iterations. We have to check whether the computation result has become stable. From this naturally follows the condition when the computation is to be terminated:

$$\xi^{(i)} = \pi S(\Gamma) - \sum_{n=1}^{k} \sum_{D \in \mathcal{L}_n^{(N-i)}}^{|\mathcal{L}_n^{(N-i)}|} \#(\Gamma_n \cap D)$$

$$\sqrt{\frac{1}{m-1} \sum_{n=0}^{m-1} (\xi^{(n)} - \overline{\xi})^2} \leqslant \varepsilon. \tag{8}$$

Here, ε is the desired precision, and $\overline{\xi} = \frac{1}{m} \sum_{n=0}^{m-1} \xi^{(n)}$ is the average, i.e., we terminate the computations when the error becomes satisfactorily small.

5 Bounding surface

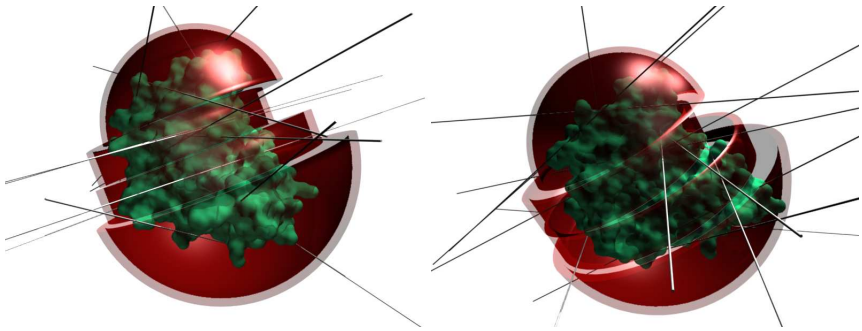


Fig. 2. Protein molecule bounded by the sliced capsule. Shown from different angles.

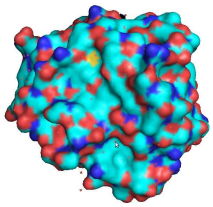
When the Cauchy–Crofton method is used to compute the area of an arbitrary given surface, it is important to understand that if we assume the area of the bounding surface to have been computed satisfactorily, we can reasonably assume that we have probably computed the area of the inscribed surface to the same satisfaction as well. Spherical bounding surfaces are usually used (as shown in the Fig. 1) to compute the

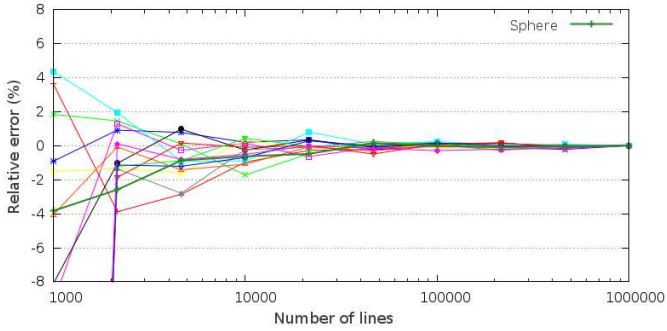
surface area when applying the Cauchy–Crofton method [2, 1] because those surfaces are simple, easily constructible, and their surface area is given by the analytical formula. Spherical surfaces have disadvantages as well. One of those disadvantages is that they poorly approximate the topology of a given surface.

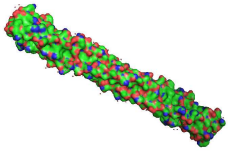
Here we show a different bounding surface that fits into the framework of the generalized Cauchy–Crofton method (5).

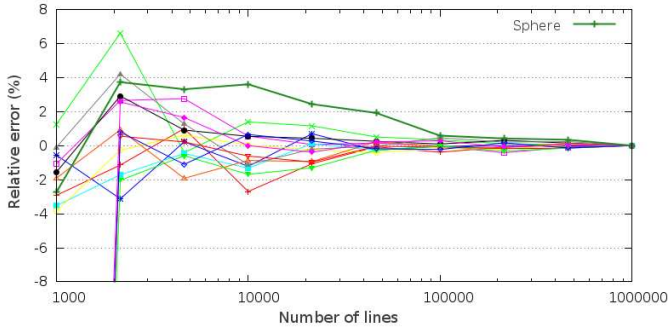
6 Test results

We have tested the generalized Cauchy–Crofton method with protein data by computing their surface area using two schemes: a spherical bounding surface and sliced capsules with the number of slices $3 \leqslant k \leqslant 15$.









As we can see from the figures above, the surface area computation is greatly influenced by a bounding surface. The speed of convergence depends on how well a bounding surface approximates the topology of the examined surface. The first case shows that for proteins whose topology is well approximated by a spherical surface both schemes converge at comparable speed. The second case emphasizes that a spherical bounding surface poorly approximates the examined surface, and the partitioned scheme offers better numerical characteristics.

References

[1] D. Juba and A. Varshney. Parallel, stochastic measurement of molecular surface area. *Journal of Molecular Graphics and Modelling*, **27**(1):82–87, 2008.

[2] X.Li, W. Wang, R.R. Martin and A. Bowyer. Using low-discrepancy sequences and the crofton formula to compute surface areas of geometric models. *Computer-Aided Design*, **35**(9):771–782, 2003.

REZIUMĖ

Gaubiančiojo paviršiaus įtaka Koši–Kroftono metodo tikslumui

J.V. Daugmaudis, A. Laurynėnas, F. Ivanauskas

Šiame darbe pasiūlėme ir realizavome geresnėmis, nei anksčiau aprašytų panašių algoritmų, skaitinėmis charakteristikomis pasižymintį metodą (izo)paviršių plotui skaičiuoti. Aprašomas stochastinis ploto matavimo algoritmas buvo išbandytas su geometriniais objektais bei su baltymų modeliais. Vienas iš aprašomo metodo privalumų yra tas, kad kiekvienos tiesės sankirtos taškų su tiriamu paviršiumi skaičius gali būti nustatomas lygiagrečiai, nepriklausomai nuo kitų imtyje esančių tiesių. Šita algoritmo savybė leidžia ypatingai efektyviai išnaudoti šiuolaikinių daugiabranduolinių procesorių skaičiavimo pajėgumus ir pritaikyti algoritmą skaičiavimams klasteryje atlikti.

Raktiniai žodžiai: Koši–Kroftono metodas, gaubiantieji paviršiai, molekulės paviršiaus ploto matavimas, stochastinis algoritmas