

Numeric Model of N -Body Problem on Domains*

Chao Liu, Shesheng Zhang[#]

School of Science, Wuhan University of Technology, Wuhan
Email: [#]sheshengz@qq.com

Received: Oct. 31st, 2012; revised: Nov. 27th, 2012; accepted: Dec. 8th, 2012

Abstract: In this paper, proximately compute interacted N -body problem by analytic function, Analytic expression of radial distribution function of particles interacting is denoted by multiple integral, the analysis solution is obtained on double sphere domains. In the case of $M = 1800$, the error of analytic solution is less than 0.001, and numeric CPU time is much less than direct computation. The two polypeptide chains protein radial function is obtained, and points out that this method needs to be improved.

Keywords: Particle Simulation; N -Body Problem; Radial Distribution Function; Multiple Integral

N 体问题复合区域解析函数近似计算*

刘 超, 章社生[#]

武汉理工大学理学院, 武汉
Email: [#]sheshengz@qq.com

收稿日期: 2012 年 10 月 31 日; 修回日期: 2012 年 11 月 27 日; 录用日期: 2012 年 12 月 8 日

摘 要: 研究粒子模拟的计算问题, 推导出单连通区域和复连通区域的粒子相互作用径向分布函数的解析解, 得到了双球域的分析解, 对计算复杂性进行了研究, 结果表明, 当粒子数 $M = 1800$ 时, 计算误差小于千分之一, 分析解的计算时间非常少。对蛋白质的距离函数计算结果表明, 分析解能显著减少计算量。

关键词: 粒子模拟; N 体问题; 径向分布函数; 多重积分

1. 引言

在粒子模拟数据分析中, 双体关联函数查询 - 空间距离直方图(SDH)的研究^[1,2]极其重要。文献[3]介绍了 Barnes2Hut 算法(BHA)和快速多极子方法(FMM), 认为都是基于树结构的快速算法。BHA 可快速计算各点受到的场力, 计算复杂度为 $O(N \log N)$, 但计算精度通常只有 1%; FMM 通过层次划分和位势函数的多极子展开计算各点位势, 其复杂度为 $O(N)$, 却能达到任意精度。SDH 也是描述系统的一系列的重要物理量的主要建筑模块, 是一种被称为径向分布函数(RDF)^[4]

的连续统计分布函数的直接估计。因此, RDF 可看成是连续化的 SDH。RDF 对计算物理系统热力学特征是不可或缺的^[5], 直接与该系统的结构因子(structure factor)相关联^[6]。在实际应用中, 为简化计算, RDF 通常是以 SDH 这样一种非连续的形式出现。文献[7]讨论了 N 体问题解析函数近似计算, 用多重积分表示粒子相互作用径向分布函数的解析表达式, 求出了圆盘区域和球体区域相互距离分析解。文献[7]的工作表明, 用解析表达式计算径向函数, 能显著减少计算量和计算误差。

文献[7]中的结论是基于假设粒子分布是稠密的。但在实际中, 存在粒子分布是多区域的情况, 如原子在具有多链的蛋白质中的分布^[8], 其分布在多个区域,

*基金项目: 国家自然科学基金资助项目(No. 51139005)。

[#]通讯作者。

若用文献[7]中的方法计算, 将会产生较大误差。本文将在文献[7]的工作基础上, 建立多区域的 N 体问题解析函数近似计算模型, 用多重积分表示粒子相互作用径向分布函数的解析表达式。

2. 多区域基本计算原理

在文献[7]中, 径向分布函数(RDF)的定义为

$$P(r) = \frac{F_h(r)}{\sum F_h(r)} \quad (2.1)$$

式中 $F_h(r)$ 是在粒子之间距离为 r , 厚度为 h 的球壳内粒子距离数量, 称为距离和函数, 分母为分子所有取值的和。设粒子坐标为 $X_i, i=1, \dots, N$, 则距离和函数 $F_h(r)$ 可表示为:

$$F_h(r) = \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N (\delta_h r^* - r) \quad (2.2)$$

这里 $r^* = |X_i - X_j|$ 为 X_i 与 X_j 之间的距离, 函数 $\delta_h = 1$ (当 $|r^* - r| \leq h/2$ 时), 函数 $\delta_h = 0$ (当 $|r^* - r| > h/2$ 时)。根据空间距离直方图(SDH)的定义^[2]。为了消去重复累加, 式中除以 2。用一个立方体 Ω 包含所有粒子。按粒子分布的稀疏情况, 将立方体 Ω 划分为 N' 个子区域 $\Omega_k, k=1, \dots, N'$ 。则距离和函数 $F_h(r)$ 可表示为:

$$F_h(r) = \frac{1}{2} \sum_{i=1}^{N'} \sum_{j=1}^{N'} F_{i,j} \quad (2.3)$$

式中 $F_{i,j}$ 为第 i 个区域中的粒子与第 j 个区域中的粒子之间的距离和函数。当 $j=i$ 时, 表示第 i 个区域内部粒子之间的距离和函数。当 $j \neq i$ 时, 表示第 i 个区域中的粒子与第 j 个区域中的粒子之间的距离和函数。距离和函数 $F_h(r)$ 也可写为:

$$F_h(r) = \sum_{i=1}^{N'} F_{i,i} + \sum_{i=1}^{N'} \sum_{j=i+1}^{N'} F_{i,j}$$

下面用高维重积分近似方法计算粒子之间的距离和函数。

3. 单区域内距离和函数

当 $N'=1$ 时, 距离和函数为一个单区域, 用一个立方体 Ω 包含单区域内所有粒子, 再将立方体划分为 M 个边长为 H 的小立方体 $d\Omega$, 第 i 个小立方体 $d\Omega_i$ 内的粒子数为 $n(i)$, 小立方体的中点坐标为 Y_i , 第 i 个小立方体中点和第 j 个小立方体中点的距离为

$|Y_i - Y_j|$, 用 $n(i)n(j)|Y_i - Y_j|$ 近似表示两个小立方体内所有粒子之间的距离的和。将下标 i, j 从 1 取到 M , 然后求和, 则有 $F_h(r)$ 的近似计算式:

$$F_h(r) \approx \sum_{j=1}^M \sum_{i=1}^M n(i)n(j)\delta_h(r^* - r) \quad (3.1)$$

式中 $r^* = |Y_i - Y_j|$, 设 dV 为小立方体的体积, 记 $\rho_i = n(i)/dV$, 则上式为:

$$F_h(r) \approx \sum_{j=1}^M \rho_j dV \sum_{i=1}^M \rho_i dV \int_{-h/2}^{h/2} \delta_h(r^* - r) dr$$

式中 δ 为 delta 函数。根据多维积分的定义, 上式可写为

$$F_h(r) \approx \int_{-h/2}^{h/2} F(r) dr$$

$$F(r) = \int_{\Omega} \rho(X) dX \int_{\Omega} \rho(X') \delta(r^* - r) dX' \quad (3.2)$$

上式是 N 个粒子径向分布函数的近似积分表达式, 其积分区域为立方体区域 Ω 的乘积 $\Omega^* \Omega$, 密度 $\rho(X)$ 取对应小立方体的 $\rho_i = \rho(X_i)$ 值。上式还可化为下面形式:

$$F(r) = \int_{\Omega} \rho(X) f(X, r) dX$$

$$f(X, r) = \int_{\Omega} \rho(X') \delta(r^* - r) dX' \quad (3.3)$$

这里 $f(X, r)$ 是以 X 为圆心, r 为半径的球面与区域 Ω 相交部分密度积分值。 $F(r)$ 是径向连续分布函数。当区域 Ω 为圆盘和球体, 密度为均匀分布时, 可以得到解析表达式。

3.1. 圆盘解

设密度为均匀分布 ρ , 区域 Ω 为半径 R 的圆盘, 当 $0 < r \leq R$ 时, 用上面公式求得积分值:

$$F(r)/\rho = r\pi^2 \left[R^2 + (R-r)^2 \right] + 4r\pi \int_{R-r}^R q\alpha dq \quad (3.4)$$

当 $R < r \leq 2R$ 时,

$$F(r)/\rho = r\pi^2 \left[R^2 - (r-R)^2 \right] + 4r\pi \int_{r-R}^R q\alpha dq \quad (3.5)$$

3.2. 球域解

考虑密度为均匀分布 ρ , 区域 Ω 为半径 R 的球域。用上面计算公式, 当 $0 < r \leq R$ 时, 可求得

$$F(r)/\rho = r\pi^2 \left[\frac{16}{3}r(R-r)^3 + R^4 - 2r^2R^2 + \frac{8}{3}rR^3 - (R-r)^3 \left(R + \frac{17}{3}r \right) \right]$$

类似地, 当 $R < r \leq 2R$ 时, 有

$$\frac{F(r)}{\rho} = r\pi^2 \left[R^4 - 2r^2R^2 + \frac{8}{3}rR^3 + (r-R)^3 \left(R + \frac{r}{3} \right) \right] \quad (3.6)$$

4. 双区域径向分布

我们考虑粒子分布在两个区域 A 和 B , 区域内分别有 N 和 N' 个粒子, 对应粒子密度分别为 ρ 和 ρ' (粒子数除区域的体积), 则粒子距离和函数可以表示为

$$F_h(r) = F_{11}(r) + F_{12}(r) + F_{22}(r) \quad (4.1)$$

式中 $r = mh$, h 为步长, $m = 1, 2, 3, \dots$ 。 $F_{11}(r)$ 为区域 A 内部粒子产生的距离和函数, $F_{22}(r)$ 为区域 B 内部粒子产生的距离和函数, $F_{12}(r)$ 为区域球 A 中的粒与区域 B 的粒子相互作用产生的距离和函数。若 A, B 区域为球域, 则有区域内函数表达式, 即 $F_{11}(r)$ 和 $F_{22}(r)$ 用上节中的公式计算:

$$\begin{aligned} F_{kk}(r) &= b_k \eta_k \left[\frac{1}{2} - \eta_k^2 + \frac{4}{3} \eta_k - \frac{1}{2} \eta_{k1}^3 + \frac{1}{6} \eta_k \eta_{k1}^3 \right] \\ \eta_k &= r/R_k \quad \eta_1 = 1 - \eta_k \quad 0 < r \leq 2R_k \\ F_{kk}(r) &= 0 \quad r > 2R_k \end{aligned} \quad (4.2)$$

式中 $k = 1, 2$, 且 $b_k = \rho^2 R_k^3 \pi^2 h$, 以及 $R_1 = R, R_2 = R'$ 。下面推导 $F_{12}(r)$ 的计算公式。

设点 Y 为球 B 表面点, 坐标为 Y , 它到点 A 的距离为 L , 则点 Y 与球 A 中粒子距离和函数为:

$$Y_h(r) = \sum_{j=1}^N (\delta_h |X_j - Y| - r) \quad (4.3)$$

根据空间距离直方图(SDH)的定义^[2], 将球域 A 划分为 M' 个子区域 $d\Omega$, 第 i 个子区域 $d\Omega_i$ 内的粒子数为 $n(i)$, 子区域的中点坐标为 X_i , 第 i 个子区域中点和区域外点 Y 的距离为 $|X_i - Y|$, 用 $n(i)|X_i - Y|$ 近似表示第 i 个子区域内所有粒子与在 Y 点粒子的距离的和。将下标 i 从 1 取到 M' , 然后求和, 则有 $Y_h(r)$ 的近似计算式:

$$Y_h(r) \approx \sum_{i=1}^{M'} n(i) \delta_h (|X_i - Y| - r) \quad (4.4)$$

设 dV 为子区域的体积, 记 $\rho_i = n(i)/dV$, 应用 δ 函数,

上式可写为:

$$Y_h(r) \approx \sum_{i=1}^{M'} \int_{-h/2}^{h/2} \delta(|X_i - Y| - r) dr \rho_i dV$$

根据多维积分的定义, 当 dV 非常小时, 上式可近似写为

$$\begin{aligned} Y_h(r) &\approx \int_{-h/2}^{h/2} Y(r) dr \\ Y(r) &= \int_0 \rho(X) \delta(r - |X - Y|) dX \end{aligned} \quad (4.5)$$

这里 $Y(r)$ 为点 Y 与球 A 中的粒子距离为 r 的个数, 称为点 Y 到球 A 的距离密度函数。当粒子密度 ρ 和 ρ' 都为常数时, 由点与球的几何关系, 经过积分运算, 有解析计算式:

$$\begin{aligned} Y(r) &= \rho \frac{r\pi}{L} \left[R^2 - (L-r)^2 \right] \\ L-R &\leq r \leq L+R \end{aligned} \quad (4.6)$$

当 $r < L-R$ 或 $r > L+R$ 时, $Y(r) = 0$ 。上式只与距离 L 和 R 有关, 与粒子坐标 Y 无关。容易求得, 球心在 A 点, 半径为 L 的球与球 B 的相交面积为

$$S = \frac{L\pi}{a} \left[R^2 - (a-L)^2 \right] \quad a-R' \leq L \leq a+R'$$

上面两式相乘, 计算得球 B 内与球 A 的球心距离为 L 的全部粒子与球 A 中的粒子的距离密度函数为:

$$g(L, r) = \rho\rho' \frac{r\pi^2}{a} \left[R^2 - (L-r)^2 \right] \left[R^2 - (a-L)^2 \right] \quad (4.7)$$

上式对 L 从 L_1 到 L_2 积分, 计算得两球距离中粒子相互距离为 r 的距离密度函数:

$$f_{12}(r) = G(L_2, r) - G(L_1, r) \quad (4.8)$$

式中

$$\begin{aligned} G(L, r) &= \rho\rho' \frac{r\pi^2}{3a} \left\{ R^2 \left[3R^2L - (L-a)^3 \right] \right. \\ &\quad \left. + \left[(r-a)^2 - R^2 \right] (L-r)^3 \right. \\ &\quad \left. + \frac{3}{10} (3r+2L-5a)(L-r)^4 \right\} \end{aligned}$$

且有 L_1 和 L_2 取值计算式

$$L_1 = \begin{cases} r-R & r-R > a-R' \\ a-R' & r-R \leq a-R' \end{cases}$$

$$L_2 = \begin{cases} r+R & r+R \leq a+R' \\ a+R' & r+R > a+R' \end{cases}$$

则有距离和函数 $F_{12}(r)$ 的计算公式

$$F_{12}(r) = \int_{-h/2}^{h/2} f_{12}(r) dr = \int_{-h/2}^{h/2} [G(L_2, r) - G(L_1, r)] dr \quad (4.9)$$

且有: $-R-R' < r-a < R+R'$, 当 $R=R'$ 时, 有退化表达式

$$\begin{aligned} L_1 &= a-R & L_2 &= r+R & r &\leq a \\ L_1 &= r-R & L_2 &= a+R & r &> a \end{aligned}$$

上面公式只与球半径和两球心之间的距离有关, 与球内粒子密度有关, 与粒子数 N 无直接关系。因而, 用上面公式计算距离和函数的计算量不随粒子数 N 的增加而增加。

5. 计算复杂性

例 1: 设两球半径 $R=R'=1$, 球心相距 $a=3$, 每个球内随机均匀分布 M 个粒子。研究粒子的分布。

本例中两球相距大于球的半径, 它们能放置于半径为 2.5 的大球中。显然, 用半径为 2.5 的大球的粒子径向分布解析公式计算将带来较大的误差。本文将用上面双球粒子径向分布解析公式计算。

由于半径 $R=1$, 则有

$$G(L, r) = \rho \rho' \frac{r \pi^2}{3a} \left\{ 3L - (L-a)^3 + [(r-a)^2 - 1](L-r)^3 + \frac{3}{10}(3r+2L-5a)(L-r)^4 \right\} \quad (5.1)$$

有 r 取值范围: $1 < r < 5$; 且有

$$L_1 = \begin{cases} r-1 & r > a \\ a-1 & r \leq a \end{cases} \quad L_2 = \begin{cases} r+1 & r \leq a \\ a+1 & r > a \end{cases}$$

此时, 只有两种情况: 1) $r \leq 3$, $(L_1, L_2) = (2, r+1)$; 2) $r > 3$, $(L_1, L_2) = (r-1, 4)$ 。

两球相互作用的距离和函数 $F_{12}(r)$ 的用近似计算公式

$$F_{12}(r) = [G(L_2, r) - G(L_1, r)]h \quad (5.2)$$

设用直接方法求出距离为 r 落入区间 $[mh-h, mh]$ 的粒子数量为 $F_{hd}(r)$, 然后用上面公式求出理论对应值 $F_h(r)$:

$$F_h(r) = F_{11}(r) + F_{12}(r) + F_{22}(r)$$

式中

$$\begin{aligned} F_{kk}(r) &= \pi^2 \rho^2 h r \left[\frac{1}{2} - r^2 + \frac{4}{3}r - \frac{1}{2}r_1^3 + \frac{1}{6}rr_1^3 \right] \\ r_1 &= 1-r \quad 0 < r \leq 2 \\ F_{kk}(r) &= 0 \quad r > 2 \end{aligned} \quad (5.3)$$

取粒子数 $M=500$, 半径 $R=R'=1$, 两球心相距离 $a=3R=3$, 步长 $h=0.1$; 球中粒子点用随机方法产生。

定义平均误差为 Er

$$Er = \frac{1}{K} \sum_{k=1}^K \left| \frac{F_h(kh)}{\sum_{k=1}^K F_h(kh)} - \frac{F_{dh}(kh)}{\sum_{k=1}^K F_{dh}(kh)} \right| \quad (5.4)$$

式中 Er 为距离 r 的分割区间数。当 $R_1=1$; $R_2=1$; $a=3$, $h=0.1$ 时, 误差 Er 和直接法 CPU 时间 T 随粒子数 M 的变化见表 1。表中 M 为球内粒子数, 计算误差 Er 乘 1000。计算时间为直接法的 CPU 计算时间。由表可知, 当非常少的粒子数 $M=100$ 时, 误差小于千分之二。当粒子数增加后, 误差总体上减少, 但发现有波动, 原因是粒子点是随机选取的, 每计算一次, 粒子坐标都随机改变。表中给出了直接法计算 CPU 时间, 随着粒子数增加, 计算 CPU 时间增加很快。而用公式进行的理论计算, 计算 CPU 时间几乎为零。

Table 1. The Error and CPU time T varied with number M
表 1. 误差 Er 和直接法 CPU 时间 T 随粒子数 M 的变化

粒子数 M	$Er \times 1000$	时间 T 秒
100	1.4313	0.407
200	1.2107	0.438
400	1.0989	1.265
700	0.72825	3.547
1000	0.89171	7.125
1200	0.76292	10.078
1500	0.73498	16.031
1800	0.70194	22.531
2000	0.64909	27.86
3000	0.73872	62.828
4000	0.69721	111.7

6. 生物中的应用

径向函数在蛋白质设计中有重要的地位, 快速计算径向函数能显著提高设计速度^[8]。应用上面理论, 我们对蛋白质原子间距离径向分布进行计算, 选取蛋白质 1B2M, 在蛋白质数据库 PDB 中下载文件 1B2M.pdb, 去掉与蛋白质无关的数据, 保留二条蛋白质肽链, 它的第一个肽链有 781 个原子, 全部原子质心在点 $A(13.239, 17.966, 3.696)$, 坐标用原子单位, 下面有关数据都使用原子单位。第二个链有 782 个原子, 全部原子质心在点 $B(1.9036, 39.84, 19.184)$, 两质心相距 29.05 原子单位。以质心 A 和 B 作半径为 $R = 14.5$ 的球, 分别包含有两个 603 个原子和 605 个原子。作为本文理论在蛋白质中的初步应用, 本例不计算球外粒子对径向函数的贡献, 不考虑原子质量, 只将原子当成一个粒子。原子数除体积, 得近似粒子密度分别为 0.04722 和 0.047376, 用前面理论进行了径向和函数计算, 并与直接法计算结果进行了比较, 结果表明, 计算误差为千分之 0.9839, 直接法计算 CPU 时间这 2.703 秒, 理论计算 CPU 时间几乎为零。计算结果表明, 随着半径 R 的增大, 球内原子数增加, 计算误差减少, 直接法计算 CPU 时间增加。

7. 结束语

分析解能显著减少距离函数的计算量, 提高计算

精度。本文根据子模拟的计算问题, 推导出单连通区域和复连通区域的粒子相互作用径向分布函数的解析解, 得到了双球域的分析解, 对计算复杂性进行了研究, 结果表明, 当%粒子数 $M = 1800$ 时, 计算误差小于千分之一, 分析解的计算时间非常少。对蛋白质的距离函数计算结果表明, 分析解能显著减少计算量。

参考文献 (References)

- [1] 胡旻, 祝大军, 刘大刚, 周俊, 刘盛纲. 粒子模拟软件吸收边界的研究[J]. 强激光与粒子束, 2006, 18(8): 1315-1318.
- [2] Y.-C. Tu, S. P. Chen and S. Pandit. Computing distance histograms efficiently in scientific databases. Proceedings of ICDE, Shanghai, 29 March-2 April 2009: 796-807.
- [3] 王武, 冯仰德, 迟学斌. 树结构在 N 体问题中的应用[J]. 计算应用研究, 2008, 125(1): 42-44.
- [4] M. Bamdad, S. Alavi B. Najafi and E. Keshavarzi. A new expression for radial distribution function and infinite shear modulus of Lennard-Jones fluids. Chemical Physics, 2006, 325(2-3): 554-562.
- [5] 陈念贻, 徐驰, 李通化. LiF-KCl 熔盐溶液的 Monte Carlo 法计算机模拟研究——I. 径向分布函数和热力学性质[J]. 中国科学(B 辑), 1987, 1: 21-26.
- [6] A. Filipponi. The radial distribution function probed by X-ray absorption spectroscopy. Journal of Physics: Condensed Matter, 1994, 6(41): 8415-8427.
- [7] 陈绍平, 章社生. N 体问题解析函数近似计算[J]. 数值计算与计算机应用, 2011, 2: 143-147.
- [8] F. Eshel, Y. Yang, S. S. Zhang and Y. Q. Zhou. Predicting continuous local structure and the effect of its substitution for secondary structure in fragment-free protein structure prediction. Structure, 2009, 17(11): 1515-1527.