文章编号:1001-2486(2003)05-0068-05

半隐式半 Lagrangian 时间积分及其可扩展并行算法设计*

张卫民,朱小谦,曹小林
(国防科技大学计算机学院,湖南 长沙 410073)

摘 要:目前谱模式仍然是全球数值天气预报业务模式的主流。针对全球数值天气预报谱模式,研究两个时间层的半隐式半 Lagrangian 时间积分格式以及用于计算起始点的准三次空间插值方法,提出了按需通讯的可扩展并行算法设计,在由4个双 CPU SMP 结点组成的 Linux 机群环境下,该算法的8任务相对于4任务的加速比达到了1.65,取得了良好的并行效果。

关键词:半 Lagrangian;谱模式;可扩展并行算法;Linux 机群系统
中图分类号:TP301.6 **文献标识码**:A

The Scalable Parallel Algorithm Implementation of Semi-Lagrangian and Semi-implicit Time Integration Scheme

ZHANG Wei-min, ZHU Xiao-qian, CAO Xiao-lin
(College of Computer, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: Spectral models are still the mainstream in global operation NWP now. Based on global spectral model, we discuss two-time-level semi-Lagrangian and semi-implicit integration scheme and quasi-cubic spatial interpolation for the terms evaluated at the departure point. A on-demand massage-passing scalable parallel algorithm has been developed. The algorithm is examined in a 4 nodes dual-CPU SMP Linux cluster. The relative speedup of 4 tasks to 8 tasks is 1.65.

Key words: semi-Lagrangian; spectral model; scalable parallel algorithm; Linux cluster

所谓半隐式格式,是指在同一方程中,对激发快波的项用隐式表示,对描述慢波的项用显示表示所构成的差分格式。由于激发快波的项有时不全是线性的,为了求解方便,只将其中的线性部分用隐式表示。而半 Lagrangian 方法是指在整个积分过程中不对同一气块沿其路径追踪,而是在同一时间步长内追踪终点总是在网格点上的气块。半 Lagrangian 方法与 Euler 方法相比具有较好的计算稳定性和较高的计算精度,从而可以加长积分的时间步长,提高数值积分的效率。如欧洲中期数值天气预报中心(ECMWF)的 T213L31 模式,若采用 Euler 方法求解,其时间步长一般要求取为 3min,而若采用三个时间层的半 Lagrangian 格式则时间步长可取为 15min,若采用二个时间层的半 Lagrangian 格式进一步可取为 30min,而两者进行 10 天的预报结果几乎是一致的。

在数值天气预报中的半 Lagrangian 方法的发展始于20世纪80年代,Robert[1] 首先将半 Lagrangian 方案与半隐式方案结合起来,提出了三个时间层的半隐式半 Lagrangian 方案,并论证了这种方法的时效优势。McDonald[3] 等提出了两个时间层的半隐式半 Lagrangian 方案,进一步提高了半 Lagrangian 方案的数值积分效率。ECMWF 于 1991 年在业务系统中实现了三个时间层的半 Lagrangian 谱模式,1996 年进一步实现了两个时间层的半 Lagrangian 谱模式[2],目前时间积分采用半隐式半 Lagrangian 的谱模式是各天气预报中心的全球中期预报业务模式的主流。

分布式并行计算的实质是,利用多个处理机通过数据通信协调完成单一计算任务。在并行计算机上实现并行计算的关键是将计算任务划分为多个独立的子任务。半 Lagrangian 方法由于需要计算各网格点在上一个时间步的起始点和时间步中间点处的各种量值,因此半 Lagrangian 谱模式相对于基于 Euler

* 收稿日期:2003-03-15
基金项目:国家自然科学基金资助项目(40245023)
作者简介:张卫民(1966—),男,副研究员,硕士。

描述的谱模式在并行计算实现方面有更大的困难性。ECMWF 的 Dent[4] 等利用区域分裂研究半隐式半 Lagrangian 谱模式的并行计算,并首先在基于共享主存 CRAY Y-MP 上实现了多任务并行业务系统,之后 Dent[5] 等引入了虚拟边界区 halo 的概念研究分布并行计算的实现方法,于 1996 年在 Fujitsu 的 VPP700 向量并行计算机上实现了 IFS 分布并行版本的业务化。由于 halo 方法的可扩展性差,半隐式半 Lagrangian 方法的可扩展并行算法是有待进一步深入研究的问题。

1 两个时间层的半 Lagrangian 时间积分格式

描述大气运动模式的控制方程可写成如下一般的微分方程形式:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = R = L + N$$

其中,L 为 R 的线性部分,N 为 R 的非线性部分。二个时间层的半 Lagrangian 方法的关键思想是根据 t_n 和 $t_n - \Delta t$ 时刻的已知风速外推出 $t_n + \Delta t/2$ 时刻的风速。利用这些风速可以计算出在 $t_n + \Delta t$ 时刻到达网格点在 t_n 时刻的起始点。然后利用 t_n 时刻的变量值通过插值得到起始点处的变量值,最后计算出 $t_n + \Delta t$ 时刻到达点处的变量值,两个时间层的半 Lagrangian 时间积分方案可以写成以下形式:

$$\frac{X_A^+ - X_D^-}{\Delta t} = \frac{1}{2}(L_D^- + L_A^+) + \frac{1}{2}(N_D^* + N_A^*)$$

其中 $X_A^+ = X(x, t_n + \Delta t)$ 为到达点处在 $t_n + \Delta t$ 时刻的值;$X_D^- = X(x - \alpha, t_n)$ 为出发点处在 t_n 时刻的值;L_D^- 和 L_A^+ 为相应点的线性项;N^* 为通过外推法得到的 $t_n + \frac{1}{2}\Delta t$ 非线性项:

$$N^* = \frac{3}{2}N(t_n) - \frac{1}{2}N(t_n - \Delta t)$$

α 为 Δt 时间内质点运动经过的路径长度,可通过迭代方法求解位移方程得到:

$$\alpha^{k+1} = \Delta t V^*(x - \frac{1}{2}\alpha^k, t_n + \frac{1}{2}\Delta t)$$

$$\alpha^0 = \Delta t V^*(x, t_n + \frac{1}{2}\Delta t)$$

三维风场 $V^*(x, t_n + \frac{1}{2}\Delta t)$ 由时间外推求出:

$$V^*(x, t_n + \frac{1}{2}\Delta t) = \frac{3}{2}V(x, t_n) - \frac{1}{2}V(x, t_n - \Delta t)$$

两个时间层的半 Lagrangian 时间积分方案不需要进行时间滤波,其变量的选取、插值格式、谱空间的处理与通常的三个时间层的半 Lagrangian 时间积分方案一致,模式积分的总计算量进一步减少。

从上面的叙述可以看出,两个时间层半 Lagrangian 的计算主要包括两个部分:(1) 沿着质点的运动轨迹反向找出每个网格点的起始点和中点;(2) 通过插值计算起始点和中间点处的各种量值。半 Lagrangian 计算是在物理过程计算之前的格点空间中进行。为了完成计算,每个格点都需要从邻近格点得到一些信息。对 TL213L31 模式,半 Lagrangian 计算约占模式总运行时间的 25% 左右。

2 起始点量值的准三次插值方法

半 Lagrangian 时间积分方案的一个主要计算过程是利用插值方法计算起始点和中间点处的各种量值,插值计算在半 Lagrangian 时间积分方案占有很大的计算量,而且它对并行算法的设计有较大影响,为此研究适合于半 Lagrangian 时间积分方案的一种高效插值方法。McDonald[3] 曾用理论证明,求位移时对风的插值精度可比求物理量的插值精度低一阶。Staniforth[6] 也认为,当用三次插值求物理量时,用线性插值求位移已经足够。为了从半 Lagrangian 积分方案中获得满足精度要求的结果,需要仔细选择插值阶数。一般地,在轨迹的中点处,可采用线性插值估计变量值,但在起始点处必须采用三次空间插值。三次空间插值是特别费时,下面研究一种称为准三次插值的方法,它可以得到与三次空间插值方法相同精度的结果,但计算量将会有极大减少。

该方法的基础是假设网格点之间的相关性与距离呈某种正比关系，对于要考虑的连续的平流场，这个假设显然是成立的，也就是说距离较远的网格点对起始点的影响相对较小。这样在三维空间网格中，对于这些相关性较小的网格点，采用简单的线性插值计算甚至不予考虑，这样就可以适当地减少计算量，同时不会带来过多的精度损失。

假设起始点的位置坐标是 $O(X_i+\alpha, Y_j+\beta, Z_k+\gamma)$，并且网格点在各个方向上是等距的。利用图1所示的32个网格点的变量值通过插值计算起始点 O 的变量值 R。

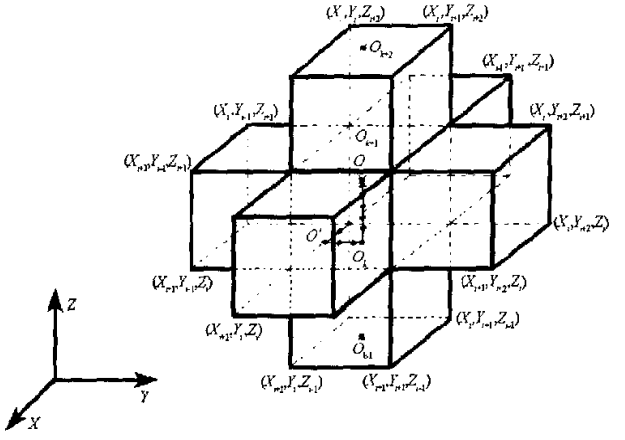

图1 计算起始点的准三次插值方法

Fig.1 Ousi-cubic spatial interpolation for the terms evaluated at the departure point

我们称 Z 坐标相同的网格点组成一个 Z 面，则上游点 O 在 $Z_{k-1}, Z_k, Z_{k+1}, Z_{k+2}$ 面上的投影分别为 $O_{k-1}(X_i+\alpha, Y_j+\beta, Z_{k-1})$、$O_k(X_i+\alpha, Y_j+\beta, Z_k)$、$O_{k+1}(X_i+\alpha, Y_j+\beta, Z_{k+1})$、$O_{k+2}(X_i+\alpha, Y_j+\beta, Z_{k+2})$。

具体的计算过程是：首先在 $Z_{k-1}, Z_k, Z_{k+1}, Z_{k+2}$ 面上分别插值求出4个投影点上的变量值：$\Psi(X_i+\alpha, Y_j+\beta, Z_{k-1}, t)$、$\Psi(X_i+\alpha, Y_j+\beta, Z_k, t)$、$\Psi(X_i+\alpha, Y_j+\beta, Z_{k+1}, t)$、$\Psi(X_i+\alpha, Y_j+\beta, Z_{k+2}, t)$。方法是采用双线性插值求解 $\Psi(X_i+\alpha, Y_j+\beta, Z_{k-1}, t)$ 和 $\Psi(X_i+\alpha, Y_j+\beta, Z_{k+2}, t)$，采用12点的三次插值求解 $\Psi(X_i+\alpha, Y_j+\beta, Z_k, t)$ 和 $\Psi(X_i+\alpha, Y_j+\beta, Z_{k+1}, t)$，然后对4个投影点的变量值做三次插值，从而得到起始点的变量值：

$$
\begin{aligned}
&\Psi(X_i+\alpha, Y_j+\beta, Z_k+\gamma, t)\\
&= \Psi(X_i+\alpha, Y_j+\beta, Z_{k-1}, t)\frac{(-\gamma)(\Delta Z-\gamma)(2\Delta Z-\gamma)}{(\Delta Z)(2\Delta Z)(3\Delta Z)}\\
&\quad + \Psi(X_i+\alpha, Y_j+\beta, Z_k, t)\frac{(-\gamma-\Delta Z)(\Delta Z-\gamma)(2\Delta Z-\gamma)}{(-\Delta Z)(\Delta Z)(2\Delta Z)}\\
&\quad + \Psi(X_i+\alpha, Y_j+\beta, Z_{k+1}, t)\frac{(-\gamma-\Delta Z)(-\gamma)(2\Delta Z-\gamma)}{(-2\Delta Z)(-\Delta Z)(\Delta Z)}\\
&\quad + \Psi(X_i+\alpha, Y_j+\beta, Z_{k+2}, t)\frac{(-\gamma-\Delta Z)(-\gamma)(\Delta Z-\gamma)}{(-3\Delta Z)(-2\Delta Z)(-\Delta Z)}
\end{aligned}
$$

在整个计算过程中，上面的准三次插值所需要的邻近点数也从64个网格点减少到32个网格点。计算量从21个三次插值减少到7个三次插值和10个线性一维插值。

对于精简高斯网格，网格不再是规则的，如果第一步插值在 λ 方向执行，那么计算量仅需增加一点。在垂直层上，当插值点位于最高两层和最低两层之间，仅需采用线性插值。超过底层和上层的外插是禁止的。

3 可扩展并行算法设计

在实现并行计算的过程中,半 Lagrangian 时间积分的两个主要部分都需要得到位于其它结点的邻近网格点,这就引入了通讯。为了方便地在分布式并行计算机上实现半 Lagrangian 并行计算,Dent[5] 等引入虚拟边界区(halo)的概念,halo 是那些围绕着分配到某个处理机的核心网格区域的邻近格点集合。halo 代表了该区域内的结点计算所需要参考的其它网格点,它一般由最大可能风速和模式时间步长所决定。引入 halo 后的一般并行实现方法是在每个时间步计算开始前交换 halo 中的所有数据。

在确定实际所需的 halo 时,考虑如何将格点分解到处理机中是十分重要的。通常的格点分解有两种方案:一种称为苹果分解,另一种称为桔子分解。在苹果分解方案中,采用连续等间隔的纬度划分格点空间,该分解只沿南北方向分解。为了满足静态的负载平衡关系,可以允许同一个纬圈上的网格点分配到两个不同的处理机上,苹果分解方案的最大并行度就是纬圈数,确定苹果分解的 halo 数据的计算相对比较简单。桔子分解对格点空间的划分采用了更为一般的形式,即不但要沿南北方向,而且要沿东西方向进行划分,分配给某个处理器上的区域是球面上的一块,而且各块之间连续。

为了高效实现半 Lagrangian 模式的可扩展并行计算,采用桔子分解的效率是比较高的,但其 halo 区域的确定相对比较复杂,我们采用了如图 2 所示的方法,将放在 *PE* 结点的核心区域所需的 halo 区域分为 6 个子块,每个子块位于不同的处理机上,形状也各不相同,halo 区域的宽度则由最大可能风速和模式时间步长两个值确定。

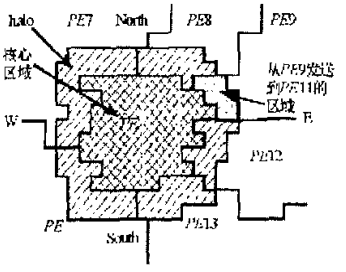

图 2 半 Lagrangian 的核心区域和 halo 区域
Fig.2 Semi-Lagrangian core and halo regions

halo 的计算需要遍历每个纬圈计算,具体算法如下:

(1) 对于本处理器核心区域,通过考虑该区域从北到南的纬圈数,计算球面上的最小和最大角度;

(2) 一个质点能在球面上运动的最大角距离(由最大风速与模式时间步长的乘积确定)被分别加减到上述的最大和最小角度上;

(3) 将角距离转换到网格点上,更多的网格点被加进来以满足插值方法的需要,插值方法越复杂,需要的点数越多;

(4) 更新 halo 的起始网格点、网格点数,以便 halo 和核心区域需要的格点数决不会大于整个纬度上的点数和插值中所需要的额外点之和。

上述 halo 方法相对比较简单,但存在两个缺点:第一,当处理器数量增加时,halo 范围并不是成比例的缩小;第二,halo 中有大量数据进行了通讯却没有被使用。

在廉价的 Linux 机群系统中,互连一般多是采用快速以太网,通讯往往是并行计算的瓶颈,如果直接采用简单的 halo 方法,并行计算效率会很低。我们知道,风可能是从区域的内部吹向边界的,这时就意味着这些边界点处的计算不需要 halo 中的值。实际实现 Lagrangian 谱模式的可扩展并行算法是根据上一时间步的风场确定区域间需要交换的数据。

4 并行试验结果

采用 4 个结点的 Linux 机群系统进行数值试验,每个结点的配置如下:主板为双 CPU 对称多处理服务器主板,CPU 为 P4 至强,内存为 1GB 的 Rambus,硬盘采用 36GB 的 Ultra SCSI/160。机群系统的操作系统采用了 RedHat Linux 7.2,4 个结点通过 100Mbps 以太网络和 KVM 网络互连。KVM 网络实现一套终端任意切换到各结点的功能,其作用是方便机群系统的管理和使用。

集群系统的消息传递并行处理软件采用了 RedHat Linux 中自带的 LAMS,FORTRAN 编译器采用 Power group 公司的编译产品 PGF90。由于谱模式对于计算精度非常敏感,所以必须采用 64 位浮点计算,其方法是在编译命令中设置 i8 和 r8 开关。

由于TL213L31需要的内存总量接近4GB,在试验用的Linux机群环境下运行TL213L31至少需要划分出4个任务,为此我们只测试了4个任务分配到4个结点和8个任务分配到4个结点两种情况的运行时间,表1列出了这两种情况下积分24步时的运行时间。

表1 算法的并行效率

Tab.1 Parallel efficiency of the algorithm

CPU数	运行墙钟时间
4	53min 30s
8	32min 20s

8个任务相对于4个任务的加速比为1.65。从计算结果看,Linux机群系统不仅能够完成全球谱模式的计算,而且有较好的并行效率。

5 结束语

在两个时间层的半隐式半Lagrangian谱模式中采用准三次插值计算起始点的各种量值在精度上可以满足要求,但在计算时所需要的邻近点数也从64个网格点减少到32个网格点,计算量从21个三次插值减少到7个三次插值和10个线性一维插值。采用基于halo区域实现的按需通讯可扩展并行算法在100Mbps以太网络连接的Linux机群系统中有较好的加速比。

参考文献:

[1] Robert A. A Semi-Lagrangian and Semi-implicit Numerical Integration Scheme for Primitive Meteorological Equations[J], J. Meteor. Soc. Japan, 1982,60: 319 - 325.

[2] White P. IFS Documentation, Part Ⅲ: Dynamics and Numerical Procedures (CY21R4)[R]. http://www. ecmwf. int/ifsdocs/Technical/index.html,Feb. 29, 2000.

[3] McDonald A. A Semi-Lagrangian and Semi-implicit Two Time Level Integration Scheme[J]. Mon. Weather Rev., 1986,114:824 - 830.

[4] Barros S, Dent D,et al. The IFS Model: A Parallel Production Weather Code[J]. Parallel Computing. 21: 1621 - 1638.

[5] Dent D, Mozdzynski. ECMWF Operational Forecasting on a Distributed Memory Platform: Forecast Model[C]. Proceeding of 7th ECMWF Workshop on the Use of Parallel Processors in Meteorology, December 2 - 6,1996.

[6] Staniforth A, Côté J. Semi-Lagrangian Integration Schemes for Atmospheric Models -- A Review[R]. Mon. Weather Rev,1991,19:2206 - 2223.

[7] McDonald A, Haugen J E. A Two Time-level, Three-dimensional Semi-Lagrangian, Semi-implicit, Limited-area Gridpoint Model of the Primitive Equations[R]. Mon. Weather Rev., 1992, 120:2603 - 2621.