

# Optimization Techniques for Adjoint Sensitivity Computation in Variational Data Assimilation

Xiaoqun Cao<sup>1</sup>, Yanling Zhao<sup>2</sup>, Yuyang Quan<sup>3</sup>, Bainian Liu<sup>1</sup>, Ya'nan Guo<sup>1</sup>

<sup>1</sup>College of Meteorology and Oceanography, National University of Defense Technology, Changsha Hunan

<sup>2</sup>The PLA 31010 Units, Beijing

<sup>3</sup>NO.92962 Troops of Chinese People's Liberation Army, Guangzhou Guangdong

Email: caoxiaoqun@nudt.edu.cn

Received: Jul. 20<sup>th</sup>, 2020; accepted: Aug. 3<sup>rd</sup>, 2020; published: Aug. 10<sup>th</sup>, 2020

---

## Abstract

In order to solve the problems of complex and difficult implementation for calculating gradients of cost function in variational data assimilation, optimization techniques for the discrete adjoint sensitivity calculation are studied. Firstly, by comparing with the direct method of gradient computation, the important role of adjoint equation in variational data assimilation is illustrated. Through theoretical derivation and detailed analysis, it is shown that the adjoint sensitivity calculation method has significant advantages in terms of computational efficiency when the dimension of the analysis vector is very large. Secondly, because the Jacobian matrix is an important factor in the adjoint sensitivity calculation, the sparse and non-zero element distribution characteristics of it are analyzed in combination with the numerical discrete schemes in atmospheric prediction models. Thirdly, a compression strategy is given to store Jacobian matrix in variational data assimilation using its sparsity. Finally, different directed acyclic graphs are used to represent the function calculation, its tangent linear and adjoint calculation processes. A fast algorithm is proposed based on graph vertex elimination to accelerate adjoint sensitivity computation demonstrated by an example. By using the graph vertex elimination method, the hotspot programs called during the minimization of the four-dimensional variational data assimilation can be optimized and improved, which will greatly reduce the cost of computation.

## Keywords

Data Assimilation, Discrete Adjoint Method, Optimization, Sensitivity Computation, Jacobian Matrix

---

# 变分资料同化中伴随敏感度计算优化技术

曹小群<sup>1</sup>, 赵艳玲<sup>2</sup>, 全育阳<sup>3</sup>, 刘柏年<sup>1</sup>, 郭亚楠<sup>1</sup>

<sup>1</sup>国防科技大学气象海洋学院, 湖南 长沙

<sup>2</sup>中国人民解放军31010部队, 北京

<sup>3</sup>中国人民解放军92962部队, 广东 广州

Email: caoxiaoqun@nudt.edu.cn

收稿日期: 2020年7月20日; 录用日期: 2020年8月3日; 发布日期: 2020年8月10日

## 摘要

针对变分资料同化中目标泛函梯度计算复杂和实现困难等问题, 研究了离散伴随敏感度计算优化技术。首先通过与梯度直接计算方法的比较, 说明了伴随方程在变分资料同化中的重要作用, 通过理论推导和分析展示了当分析变量维数特别巨大时伴随敏感度计算方法在计算效率方面的显著优势; 其次考虑到在伴随敏感度计算中Jacobian矩阵是重要影响因素, 结合大气预报方程组离散格式分析了伴随方程中Jacobian矩阵的稀疏性及非零元素分布特征; 再次, 针对变分资料同化中Jacobian矩阵的稀疏性给出了一种压缩存储策略; 最后利用有向无环图对函数计算、切线性和伴随计算过程进行了表示, 通过示例说明了基于计算图顶点消除的伴随代码优化技术, 采用顶点消除方法对四维变分资料同化极小化过程中调用的热点伴随程序进行优化改进后, 将大大减小计算量。

## 关键词

资料同化, 离散伴随方法, 最优化, 敏感度计算, Jacobian矩阵

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

资料同化的目标是通过融合时空分布不规则的多源观测信息, 为数值预报模式提供最优初值条件[1] [2] [3] [4]。资料同化在大气、海洋、电离层、非线性动力系统和流体力学等众多领域已有广泛应用, 对其计算技术的研究具有重要价值[5] [6] [7] [8] [9]。四维变分资料同化方法(4D-Var)是目前国际上气象和海洋预报业务部门所使用的先进同化方法之一, 在提高数值预报准确度方面已取得巨大成功。4D-Var本质上是一个以地球物理流体运动方程为约束条件的大规模最优化问题, 解算结果是大气或海洋数值预报模式需要的初始条件以及其它参数(例如: 模式误差、卫星观测偏差参数等) [10] [11] [12] [13] [14]。一方面, 随着大气、海洋科学和高性能计算机的快速发展, 数值预报模式分辨率不断提高, 同化系统分析变量规模已增长到  $10^9$  量级甚至更大; 另一方面, 物理和化学过程参数化方案、各种非线性观测算子的引入增加了目标泛函的复杂性。上述原因导致在变分资料同化中敏感度信息(即目标泛函关于控制变量的梯度)计算成为一个具有挑战性的科学难题。虽然基于伴随方程的敏感度计算方法在理论上具有优势, 但是在实际具体应用中存在一系列问题。首先, 伴随方程方法主要分为两类: 连续和离散方法, 两者各有优缺点, 在变分资料同化中是选择使用连续伴随方程方法还是离散伴随方法或是两者的混合? 目前尚无确定性结论, 需要进一步研究; 其次, 目前利用自动微分工具或手工编码方法都无法开发出完美的伴随模式 [15] [16] [17]; 再次, 在构造伴随模式时常常遇到不可微或“开关”问题, 不存在统一的处理方法, 需要针对不同情况研究特殊方法[18]; 最后, 伴随模式在时间上是逆向积分的, 需要将预报模式状态正向演化量作为伴随模式的系数, 因此在伴随模式积分过程中存储和计算代价都非常巨大。

本文主要针对上面的问题,研究了变分资料同化中的伴随敏感度计算优化技术。首先通过与梯度直接计算方法的比较,说明离散伴随方程在变分资料同化中的重要价值,通过理论推导和分析证明了在分析变量维数特别巨大的情况下离散伴随敏感度计算方法在计算代价方面的显著优势;其次考虑到在敏感度计算中 Jacobian 矩阵是影响计算性能的重要因素,结合大气预报方程组离散格式分析了伴随方程中 Jacobian 矩阵的稀疏性及非零元素分布特征;最后通过示例,给出了 Jacobian 矩阵存储优化策略和基于计算图顶点消除的伴随模式优化计算技术,从而为伴随方程求解优化奠定基础,最终加速变分资料同化计算过程。

## 2. 变分资料同化

下面主要通过和梯度直接计算方法的比较,证明离散伴随方程方法在计算目标泛函敏感度(或梯度)信息方面的显著优势,从而说明伴随方程在变分资料同化中的重要作用。在数值预报的变分资料同化问题中,目标泛函  $J$  可以表示为地球物理流体(主要是大气和海洋流体,下面以大气为例)运动状态变量  $\mathbf{x}$  和分析变量  $\mathbf{u}$  的泛函[1] [2] [3] [4] [5]

$$J(\mathbf{x}, \mathbf{u}) \equiv \frac{1}{2}(\mathbf{u} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{u} - \mathbf{x}_b) + \frac{1}{2} \int_0^T (\mathbf{y}_o(t) - H[\mathbf{x}(t)])^2 dt \quad (1)$$

其中  $H$  表示非线性观测算子,  $\mathbf{u}$  和  $\mathbf{x}_b \in R^N$  分别代表初始状态和背景场向量,同时在  $[0, T]$  时间段内分布有一系列观测资料  $\mathbf{y}_o \in R^M$ 。而定量描述大气流体运动的离散偏微分方程组(Navier-Stokes 方程)在形式上可表示为

$$\mathbf{F}(\mathbf{x}(\mathbf{u}), \mathbf{u}) = 0 \quad (2)$$

其中,三维大气流体运动状态变量  $\mathbf{x} = (T, p, q, v_x, v_y, v_z)^T$  是随时间和空间变化的,公式(2)中的  $\mathbf{x}(\mathbf{u})$  表示运动状态向量是分析变量的函数,即分析变量  $\mathbf{u} = (u_1, u_2, \dots, u_N)^T$  控制了同化窗口内的大气状态轨迹的变化。一般情况下,分析变量维数远远大于观测资料数量,即  $N \gg M$ 。变分资料同化问题利用数学公式可表示为

$$\min J = J(\mathbf{x}, \mathbf{u}), \text{ s.t. } \mathbf{F}(\mathbf{x}, \mathbf{u}) = 0 \quad (3)$$

当以大气运动离散方程组(2)式为约束条件的目标泛函  $J$  达到极小值时,则得到最优分析量。对(3)式表示的最优化问题,其数值求解需要向特定的优化算法提供准确梯度向量值。为了说明离散伴随方程的重要作用及其计算敏感度的显著优势,首先给出传统的梯度直接计算方法。采用微分链式法则,目标泛函(1)式对分析变量的导数可以写为[15] [16]:

$$\frac{dJ}{du_i} = \frac{\partial J}{\partial u_i} + \frac{\partial J}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial u_i}, \quad (i=1, 2, \dots, N) \quad (4)$$

在(4)式中  $\partial J / \partial u_i$  与  $\partial J / \partial \mathbf{x}$  的计算相对较容易,而  $\partial \mathbf{x} / \partial u_i$  的计算存在较大难度。考虑数值离散后的大气运动控制方程(2)式,将其线性化得到,

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial u_i} = -\frac{\partial \mathbf{F}}{\partial u_i}, \quad (i=1, 2, \dots, N) \quad (5)$$

其中,  $\partial \mathbf{F} / \partial \mathbf{x}$  是  $N \times N$  维 Jacobian 矩阵[15] [16]。梯度直接计算方法首先需要通过求解大规模线性代数方程组(5)式后得到  $\partial \mathbf{x} / \partial u_i$ , 然后将其代入(4)式中计算梯度值。由(5)式易知,对于每个控制变量  $u_i$  都需要单独求解一个大规模线性代数方程组,最终必须求解  $N$  个线性方程组后才可以计算出目标泛函  $J$  对所有控制变量  $\mathbf{u}$  的敏感度。当  $N$  值较小时,可以利用梯度直接计算方法获得梯度向量;但是,当  $N \geq 10^3$  时,梯度直接计算方法由于计算量太大而无法实现。下面继续讨论基于离散伴随方程的敏感度计算方法。

首先, 引入离散伴随向量  $\lambda$  将(4)式中的梯度表达式改写成下面形式,

$$\frac{dJ}{du_i} = \frac{\partial J}{\partial u_i} + \lambda^T \cdot \frac{\partial F}{\partial u_i}, \quad (i=1,2,\dots,N) \quad (6)$$

而伴随向量  $\lambda$  是通过下面的离散伴随方程表达式求得,

$$\left(\frac{\partial F}{\partial x}\right)^T \lambda = -\left(\frac{\partial J}{\partial x}\right)^T, \quad (7)$$

由上面(6)和(7)两个公式可以看出, 欲求得目标泛函  $J$  对  $N$  维分析变量  $\mathbf{u}$  的导数, 只需要求解一个大规模线性代数方程组(7)式[15] [16]。剩下的唯一问题是梯度直接计算和离散伴随方法是否具有 consistency, 下面证明两种不同方法计算敏感度的结果是相同的。

首先将(5)式代入(6)式, 得到

$$\frac{dJ}{du_i} = \frac{\partial J}{\partial u_i} - \lambda^T \frac{\partial F}{\partial x} \cdot \frac{\partial x}{\partial u_i} \quad (8)$$

而(8)式可以改写为

$$\frac{dJ}{du_i} = \frac{\partial J}{\partial u_i} - \left[\left(\frac{\partial F}{\partial x}\right)^T \lambda\right]^T \cdot \frac{\partial x}{\partial u_i} \quad (9)$$

将离散伴随方程(7)式代入(9)式后得到

$$\frac{dJ}{du_i} = \frac{\partial J}{\partial u_i} + \frac{\partial J}{\partial x} \cdot \frac{\partial x}{\partial u_i} \quad (10)$$

(10)式与(4)式是完全相同的, 从而证明了对于变分资料同化中的敏感度计算问题, 分别利用离散伴随方法和梯度直接计算方法求解的结果是相同的。但前者只需要求解一个  $N \times N$  规模的线性代数方程组(伴随方程(7)式), 而后者需求解  $N$  个类似规模的线性代数方程组, 从而说明伴随方法的计算代价大大减小。(6)式和(7)式一起构成了计算目标泛函关于控制变量敏感度的离散伴随方法。主要优点是: 因为消除了目标泛函梯度对状态量的敏感度依赖关系, 而且在伴随方程中也不存在对控制变量  $\mathbf{u}$  的敏感度依赖关系, 所以在最优化过程中具有迭代计算代价不依赖于控制变量规模的优点。总之, 基于离散伴随方法计算梯度向量在理论上具有显著优势: 针对以离散偏微分方程为约束条件的变分资料同化问题, 无论控制变量数量是多么巨大, 在最优化每次迭代过程中只需要求解一套离散的大气运动控制方程和伴随方程, 即计算量在理论上是数值预报模式积分运算量的二倍。另外, 与连续伴随方法相比, 离散伴随方法具有伴随模式建立过程清晰规范、易于扩展到大气/海洋多尺度和高分辨率数值预报模式等优点, 能够有效克服连续伴随方法的某些不足。

### 3. Jacobian 矩阵结构特征分析

离散伴随方法是从离散的大气运动控制方程(2)式出发, 通过一定的处理和推导后得到离散伴随方程和目标泛函关于控制变量敏感度的计算公式, 因此其最终结果将高度依赖数值预报模式所采用的离散格式[15] [16]。从(6)式可知, 目标泛函关于控制变量的敏感度是伴随状态量  $\lambda$  的线性函数, 而伴随状态量可以通过求解大规模线性代数方程组(7)式得到。从(7)式易知, 作为变分资料同化的非线性约束条件, 离散偏微分方程  $F(\mathbf{x}(\mathbf{u}), \mathbf{u})$  的 Jacobian 矩阵  $\partial F/\partial x$  的结构特征将对伴随方程数值求解有巨大影响。为了优化伴随方程和目标泛函敏感度计算, 下面对 Jacobian 矩阵的特征进行重点分析。因为大气运动控制方程一般都属于时间演变(或发展)偏微分方程组, 所以能够依据预报时间步对状态向量和约束方程进行分离和排序, 即:

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m)^T, \quad \mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_i, \dots, \mathbf{F}_m)^T \quad (11)$$

其中  $m$  是预报时间步数量,  $\mathbf{x}_i$  和  $\mathbf{F}_i$  是第  $i$  个时间步的状态向量和非线性约束方程。Jacobian 矩阵描述了非线性约束方程关于大气运动状态向量的敏感度信息, 例如, 第  $i$  个时间步的非线性约束  $\mathbf{F}_i$  关于状态向量  $\mathbf{x}_j$  的敏感度信息分布通过 Jacobian 矩阵的子块可表示如下:

$$\mathbf{J}_{ij} = \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_j} \quad (12)$$

利用上面的表示方法, 下面将说明伴随方程(7)式中的 Jacobian 矩阵  $\partial \mathbf{F} / \partial \mathbf{x}$  具有良好的稀疏结构特征。众所周知, 大气和海洋等流体运动变化只受过过去状态值影响, 而与未来状态无关, 即  $\mathbf{F}_i$  只依赖于第  $i$  个时间步之前状态向量的部分元素。换言之, 当  $j > i$  时, 所有的 Jacobian 矩阵块都是零矩阵块, 即  $\mathbf{J}_{ij} = 0$ 。因此完整 Jacobian 矩阵呈现出下面的块下三角形式。

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{J}_{21} & \mathbf{J}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{J}_{i1} & \mathbf{J}_{i2} & \vdots & \mathbf{J}_{ii} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \mathbf{0} \\ \mathbf{J}_{m1} & \mathbf{J}_{m2} & \vdots & \vdots & \mathbf{J}_{mm-1} & \mathbf{J}_{mm} \end{bmatrix} \quad (13)$$

将(13)式代入(7)式后, 则离散伴随方程将转化成下面的新线性代数方程组:

$$\begin{bmatrix} \mathbf{J}_{11}^T & \mathbf{J}_{21}^T & \vdots & \mathbf{J}_{i1}^T & \vdots & \mathbf{J}_{m1}^T \\ \mathbf{0} & \mathbf{J}_{22}^T & \vdots & \mathbf{J}_{i2}^T & \vdots & \mathbf{J}_{m2}^T \\ \mathbf{0} & \mathbf{0} & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{ii}^T & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots & \mathbf{J}_{mm-1}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{mm}^T \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_i \\ \vdots \\ \lambda_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix} \quad (14)$$

其中  $\mathbf{b}_i = -(\partial \mathbf{J} / \partial \mathbf{x}_i)^T$ 。在(14)式中, 依据不同时间步将离散伴随状态向量  $\lambda$  分割成  $m$  个部分, 即  $\lambda_i$  表示第  $i$  个时间步的伴随状态向量。根据所采用的离散格式,  $\mathbf{F}_i$  只依赖于第  $i$  个时间步附近状态变量  $\mathbf{x}$  的部分元素, 只有当时间步  $j$  在时间步  $i$  附近时才有  $\mathbf{J}_{ij} \neq 0$ , 即在(14)式中 Jacobian 矩阵上三角部分中将有许多零矩阵块。在大多数情形下, Jacobian 矩阵一般呈现块带状, 具有依赖于时间离散格式的块带宽值。例如, 当采用单时间步格式离散大气运动控制方程时, 那么  $\mathbf{F}_i$  只依赖于  $\mathbf{x}_i$  和  $\mathbf{x}_{i-1}$ , 在此情形下 Jacobian 矩阵的块带宽值是 2; 类似地, 如果采用双时间步离散格式, 则块带宽值将是 3, 等等。特别地, 如果采用显式时间差分格式离散大气运动控制方程, 则有  $\mathbf{F}_i = \mathbf{x}_i - \mathbf{f}_i(\mathbf{x}_{i-1}, \dots)$ , 那么 Jacobian 矩阵的对角块将是单位矩阵:

$$\mathbf{J}_{ii} = \frac{\partial \mathbf{F}_i}{\partial \mathbf{x}_i} = \mathbf{I} \quad (15)$$

在此情形下, 完整 Jacobian 矩阵不但是块下三角结构, 而且呈现对角阵结构。

通过上面的分析可知, 如果采用局部化时间离散格式构造数值预报模式, 那么伴随方程中的 Jacobian 矩阵将是块对角阵且具有固定块带宽值。下面通过对空间差分格式的分析, 将进一步说明每个 Jacobian 矩阵块也具有稀疏特征。在对大气运动控制方程进行数值离散时, 通常所有空间差分格式都是局部性的, 即每个网格点上的非线性约束方程  $\mathbf{F}$  只依赖于该网格点附近的的状态量。表示为:

$$\mathbf{x}_i = (\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \dots, \mathbf{x}_{k,i}, \dots, \mathbf{x}_{n,i})^T, \quad \mathbf{F}_i = (\mathbf{F}_{1,i}, \mathbf{F}_{2,i}, \dots, \mathbf{F}_{k,i}, \dots, \mathbf{F}_{n,i})^T \quad (16)$$

其中  $\mathbf{F}_{k,i}$  和  $\mathbf{x}_{k,i}$  是网格点  $k$  和时间步  $i$  处的非线性约束方程和状态向量。因为采用的是局部化空间差分格式，所以只有当格点  $j$  在格点  $k$  附近时才有  $\partial \mathbf{F}_{k,i} / \partial \mathbf{x}_{j,i} \neq 0$ ，所以即使是 Jacobian 矩阵中  $n \times n$  规模的非零矩阵块

$$\mathbf{J}_{ii} = \begin{bmatrix} \frac{\partial \mathbf{F}_{1,i}}{\partial \mathbf{x}_{1,i}} & \dots & \frac{\partial \mathbf{F}_{1,i}}{\partial \mathbf{x}_{k,i}} & \dots & \frac{\partial \mathbf{F}_{1,i}}{\partial \mathbf{x}_{n,i}} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{1,i}} & \dots & \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{k,i}} & \dots & \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{n,i}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{F}_{n,i}}{\partial \mathbf{x}_{1,i}} & \dots & \frac{\partial \mathbf{F}_{n,i}}{\partial \mathbf{x}_{k,i}} & \dots & \frac{\partial \mathbf{F}_{n,i}}{\partial \mathbf{x}_{n,i}} \end{bmatrix} \quad (17)$$

也将具有稀疏特征，其中  $n$  是空间中网格点数量。如果采用的是空间中心差分格式，则只有

$\frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{k-1,i}}, \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{k,i}}, \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{k+1,i}} \neq 0$ ，非零矩阵块  $\mathbf{J}_{ii}$  将呈现出三对角结构，表示如下：

$$\mathbf{J}_{ii} = \begin{bmatrix} \frac{\partial \mathbf{F}_{1,i}}{\partial \mathbf{x}_{1,i}} & \frac{\partial \mathbf{F}_{1,i}}{\partial \mathbf{x}_{2,i}} & 0 & 0 & 0 \\ \frac{\partial \mathbf{F}_{2,i}}{\partial \mathbf{x}_{1,i}} & \ddots & \ddots & 0 & 0 \\ 0 & \ddots & \frac{\partial \mathbf{F}_{k,i}}{\partial \mathbf{x}_{k,i}} & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & \frac{\partial \mathbf{F}_{n-1,i}}{\partial \mathbf{x}_{n,i}} \\ 0 & 0 & 0 & \frac{\partial \mathbf{F}_{n,i}}{\partial \mathbf{x}_{n-1,i}} & \frac{\partial \mathbf{F}_{n,i}}{\partial \mathbf{x}_{n,i}} \end{bmatrix} \quad (18)$$

总之，上面通过对四维变分资料同化中离散伴随方程的详细分析，可以归纳出如下主要结论。首先离散伴随方程是一个大规模线性方程组，该线性方程组的求解高度依赖于四维变分资料同化中非线性约束方程(即数值预报模式)的 Jacobian 矩阵；其次，由于针对大气运动控制方程通常采用时间和空间局部化差分格式构造数值预报模式，所以导致 Jacobian 矩阵成为一个稀疏矩阵，从而使得离散伴随方程成为一个具有良好结构的典型稀疏线性代数方程组；最后，通过利用(14)式中的块上三角结构可以实现离散伴随方程的逐步求解策略：基于块后向代入方法首先求得  $\lambda_m$ ，然后依次计算得到  $\lambda_{m-1}$ 、 $\lambda_{m-2}$ 、 $\dots$ ，最终求得  $\lambda_1$ ，即离散伴随方程的求解与数值预报模式积分方向不同，需要采用时间逆向算法。

#### 4. Jacobian 矩阵压缩技术

在四维变分资料同化中伴随方程区别于其它线性系统的是其大规模特征[16][17]。一个数值预报模式所对应线性伴随系统的维数大小是  $N = m \times n$ ，其中  $m$  是时间预报总步数，而  $n$  与空间网格点数量(达到  $10^9$  量级甚至更大)成正比例。对于典型高分辨率数值预报中的四维变分资料同化问题，伴随方程的维数大小可能是数十万亿计。因此，相对于一般线性系统伴随方程的求解需要采用不同的策略和技术。如此巨大维数的一个直接后果是无法获得存储整个线性系统所需要的计算机内存。一方面有必要使用逐步计

算方法每次求解伴随方程的一部分,即根据时间逆向顺序和基于后面时间步( $i+1$ )的伴随状态量信息,求解前一个时间步 $i$ 的伴随方程状态量,即每次只求解一个时间步上的子线性伴随系统;另一方面,可以利用第2部分中分析得到的Jacobian矩阵良好的稀疏结构特征,实现Jacobian矩阵的压缩存储,从而减小存储空间需求。

图1给出了稀疏Jacobian矩阵的压缩示意图,从图中易知:通过适当调整非零元素的分布位置,可以实现稀疏Jacobian矩阵的压缩存储,从而减小计算机存储空间需求。因为在数值预报模式中当所采用的时间和空间离散格式固定后,其相应伴随方程的Jacobian矩阵稀疏结构就不再发生变化,例如:大规模三对角矩阵,所以可以采用固定不变的压缩存储方案。从图1的表示可知:采用压缩存储技术后,存储空间需求减小为原来的三分之一;对于具有相同稀疏结构的Jacobian矩阵,随着矩阵规模增大,存储压缩率将快速提高。因此,在四维变分资料同化系统求解中,考虑到分析变量的大规模特征和Jacobian矩阵特定不变的稀疏结构,通过采用压缩存储技术将可以大大降低存储代价。

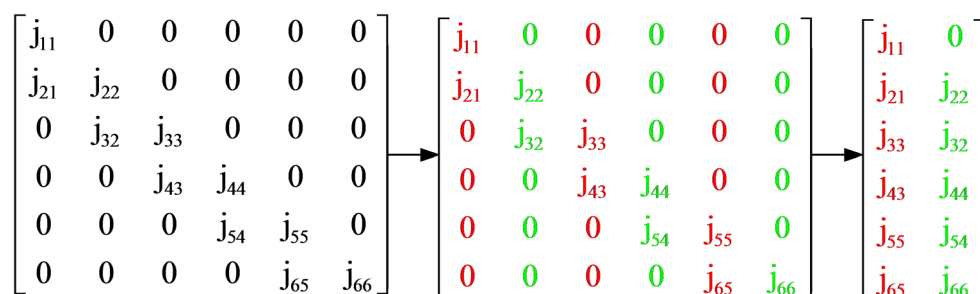


Figure 1. Diagram of Jacobian matrix compressing

图 1. Jacobian 矩阵压缩示意图

## 5. 基于计算图顶点消除的伴随计算优化技术

业务变分资料同化系统一般应用梯度类算法进行求解,因此如何快速、准确、高效地计算梯度信息,对于变分资料同化求解具有重要意义。计算梯度需要求解目标函数相对于每一个控制变量的导数。计算导数方法主要有:有限差分法、直接计算法以及伴随方法,等等。若计算目标函数相对于一个控制变量的导数,有限差分法必须至少计算两次目标函数,因此其计算量与设计变量数量成正比。此外,由于计算机有效数字位数的限制,差分步长的选择将严重影响导数的计算精度。复变量差分法可以获得目标函数相对于设计变量的精确导数,而且受差分步长影响较小,其计算目标函数相对于一个设计变量的导数需要进行一次带复变量扰动的目标函数计算,因此其计算量也与设计变量个数成正比[5]。在典型变分资料同化求解问题中,作为控制变量的初始场维数超过千万。在这种问题中,若采用有限差分法或复变量差分法,计算一次梯度将需要反复调用数值预报模式求解目标函数,产生极大的计算代价。伴随方法通过求解原始数值预报模式的伴随方程,只需要经过一次方程求解过程,即可获得目标函数相对于所有设计变量的导数。伴随方程求解梯度过程中,其计算量与控制变量个数之间基本可实现解耦合,并且精度较高,在大规模最优化问题中具有显著优势。离散伴随方法可以计算得到机器精度的目标函数导数信息,且计算代价合理,与其它方法相比具有明显优势。该方法的基本思想是在数值预报模式计算中,无论计算过程或函数如何复杂,在本质上都是执行一系列的初等计算、初等函数及逻辑运算的有限次组合。对这些初等运算迭代运用链式求导法则,然后将初等函数微分进行连续逆向累积得到复杂函数的微分信息,最终可以精确地得到目标函数的导数信息。为了说明各种方法的优劣,以计算下面函数的梯度作为示例。虽然用作实例说明的函数形式上简单,但具有典型代表性。

$$y = f(x_1, x_2, x_3) = \cos[(x_1 + x_2 + x_3) \cdot x_3^{1/2}] \quad (19)$$

首先定义独立变量  $x_i$  ( $i=1,2,3$ )和引入中间变量  $v_i$  ( $i=1,2,3,4$ ), 将函数计算过程分解成一系列的基本运算序列, 如表 1 中左栏所示。

直接计算方法是按照从自变量(Independent Variable)到因变量(Dependent Variable)的方向将求导法则应用到表 1 左栏所示的函数计算序列中。从自变量到最终函数传递中间变量关于自变量的梯度向量, 同时得到函数值以及具有机器精度的梯度值, 具体如表 1 中栏所示。离散伴随方法与直接计算方法完全相反, 是按照从因变量到自变量反方向逐次计算梯度信息, 如表 1 右栏所示, 变量  $v_i$  的伴随  $\bar{v}_i$  被定义为因变量  $y$  关于  $v_i$  的导数。在伴随方法中首先正向执行最左栏的函数计算序列, 并记录中间变量, 然后通过传递因变量与自变量之间的导数关系, 反向累积得到导数信息。

**Table 1.** Computation sequence of the function (19) and its gradient calculation by the direct and adjoint method respectively  
**表 1.** 函数(19)式的计算序列、基于直接方法和伴随方法计算梯度的序列

函数计算序列	直接计算方法	离散伴随方法
$x_1$	$\nabla x_1 = [1 \ 0 \ 0]$	$\bar{x}_i = 0, \quad i=1,2,3$
$x_2$	$\nabla x_2 = [0 \ 1 \ 0]$	$\bar{v}_i = 0, \quad i=1,2,3,4$
$x_3$	$\nabla x_3 = [0 \ 0 \ 1]$	$\bar{y} = 1$
$v_1 = x_1 + x_2$	$\nabla v_1 = \nabla x_1 + \nabla x_2$	$\bar{v}_4 = \bar{v}_4 - \bar{y} \cdot \sin v_4$
$v_2 = v_1 + x_3$	$\nabla v_2 = \nabla v_1 + \nabla x_3$	$\bar{v}_2 = \bar{v}_2 + v_3 \cdot \bar{v}_4, \quad \bar{v}_3 = \bar{v}_3 + v_2 \cdot \bar{v}_4$
$v_3 = \sqrt{x_3}$	$\nabla v_3 = \nabla x_3 / (2\sqrt{x_3})$	$\bar{x}_3 = \bar{x}_3 + \bar{v}_3 / (2\sqrt{x_3})$
$v_4 = v_2 \cdot v_3$	$\nabla v_4 = \nabla v_2 \cdot v_3 + v_2 \cdot \nabla v_3$	$\bar{v}_1 = \bar{v}_1 + \bar{v}_2, \quad \bar{x}_3 = \bar{x}_3 + \bar{v}_2$
$y = \cos v_4$	$\nabla y = -\nabla v_4 \cdot \sin v_4$	$\bar{x}_1 = \bar{x}_1 + \bar{v}_1, \quad \bar{x}_2 = \bar{x}_2 + \bar{v}_1$

图模型是在研究变量之间因果关系时被广泛应用的工具, 对表 1 中的三种类型计算都可以通过计算图表示。首先定义图  $G=(V, E)$ , 其中  $V = \{v_i | i=1,2,\dots,|V|\}$  是图  $G$  中顶点所代表的变量集合, 表示数值计算过程应用中所有计算子任务的集合;  $E = \{e_k | k=1,2,\dots,|E|\}$  是图  $G$  的边集合, 表示不同顶点代表的变量之间的依赖关系。图模型中的边表示为  $e_k = (v_i, v_j)$ , 其中  $v_i$  和  $v_j$  分别表示边  $e_k$  的起点和终点, 只有在  $v_i$  执行完毕之后才可以执行  $v_j$ 。如果  $(v_i, v_j)$  和  $(v_j, v_i)$  都属于  $E$ , 则在图  $G$  中顶点  $v_i$  和  $v_j$  之间的边为无向边, 即只是一条简单的连线。如果只有  $(v_i, v_j) \in E$ , 则在  $G$  中表现为  $v_i \rightarrow v_j$ , 即为有向边。如果  $G$  中只存在有向边或无向边, 则分别称为有向图和无向图。表 1 左栏计算过程可以通过计算图表示, 具体如图 2 所示。图中的顶点包含了函数(19)式中间变量的所有计算过程, 在计算中每一个顶点只有当它的所有前驱顶点成功执行之后才被执行。在计算图顶点集合中存在两类特殊顶点: 无任何前驱的顶点为输入顶点, 无任何后继的顶点为输出顶点。因为在图中不存在任何回路, 所以这种图一般被称为有向无环图(Directed Acyclic Graph, 简称 DAG)。如果计算出每一个顶点所定义的基本函数关于前驱顶点变量的局地偏导数, 并作为计算子任务放置在 DAG 图形边上, 那么就产生了如图 3 所示的线性化 DAG (Linearized DAG, 简称 LDAG)。从图 3 可知, 放置在图形边上的导数表达式与表 1 中间栏中的偏导数是相同的, 因此线性化 DAG 表示的是基于直接方法计算梯度的过程。因为直接方法的计算复杂性与独立变量数量成正比, 所以需要三次遍历图 3 才能计算出所有导数向量。



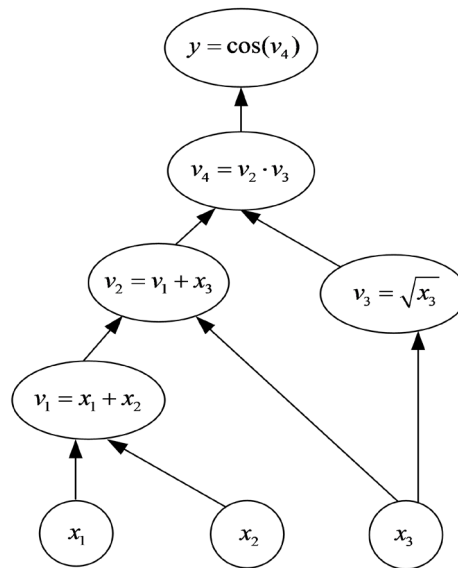


Figure 2. Directed acyclic graph of function  $f$

图 2. 函数  $f(x_1, x_2, x_3)$  的有向无环图

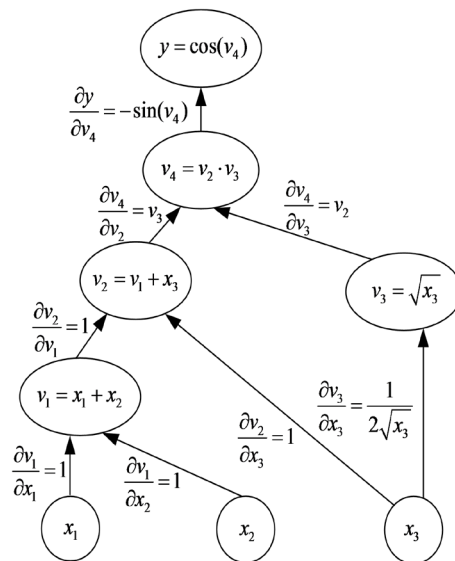


Figure 3. Linearized directed acyclic graph of function  $f$

图 3.  $f$  的线性化有向无环图

如果将图 3 线性化 DAG 中每一条有向边的方向设置为相反方向, 并根据伴随关系(表 1 右栏所示)重新设置顶点变量和每条边上的计算子任务, 则得到如图 4 所示的伴随 DAG (Adjoint DAG, 简称 ADAG)。ADAG 和 LDAG 的计算顺序完全相反, 在 ADAG 中伴随量将从依赖输出量( $y$ )逆向传播到独立输入量(即自变量  $x_1$ 、 $x_2$  和  $x_3$ ), 直到梯度被完全累积计算, 即通过对 ADAG 的一次逆向遍历计算就能获得梯度向量。从上到下依次计算这些操作产生了输入变量的伴随值( $\bar{x}_1, \bar{x}_2$ ), 即梯度量。伴随状态用零初始化。一个变量的伴随可以通过计算 LDAG 中其所有后继变量伴随量的加权和得到, 而权重就是局地导数。伴随(信息)量的传播能够通过图 4 所示的伴随 DAG 图形显示出来。伴随方法一个明显的缺点是需要存储整个计算图来完成导数计算, 在解决大规模问题时需要相当大的内存。利用稀疏分布特征, 可以实现 Jacobian

矩阵的压缩存储，从而减小计算机内存需求。下面通过采用基于计算图顶点消除方法，进一步减小伴随计算内存需求和计算代价。

对比图 4 和图 5 可知，在伴随 DAG 中消除顶点  $v_1$  和  $v_2$  后，计算梯度所需要的计算初等函数次数将由 8 次减小到 6 次，即减小的计算次数为消除的图顶点总数。说明通过消除伴随计算图中的顶点能够减小计算代价，函数(19)式表示的只是某个子计算过程或伴随程序代码的运算情形，如果对变分资料同化中每个伴随程序采用类似的方法进行改写，特别是针对某些反复调用的热点伴随程序进行优化改进，将大大减小计算量。在变分资料同化中主要计算过程是迭代最优优化过程中的切线性模式正向积分步程序和伴随模式逆向积分步程序，在采用特定的时空离散格式后两者的结构是固定不变的，因此可以通过消除计算图顶点的优化技术对程序进行改进后永久使用，而不需要每次调用之前都对伴随程序进行优化。

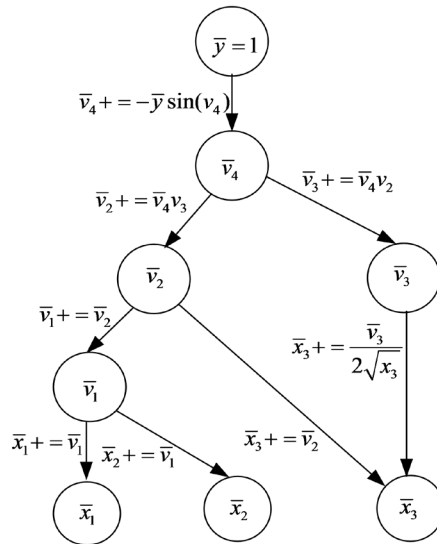


Figure 4. Adjoint directed acyclic graph of function  $f$   
图 4.  $f$  的伴随有向无环图

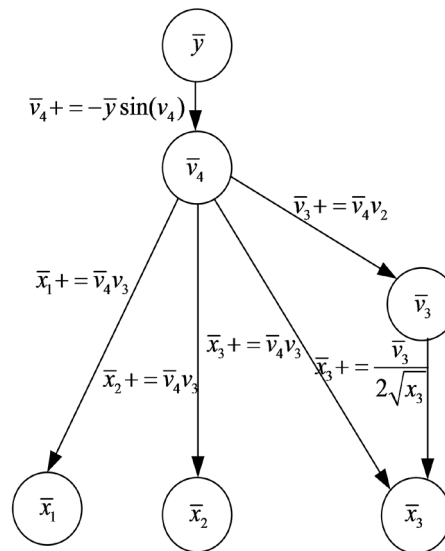


Figure 5. Improved ADAG of function  $f$   
图 5.  $f$  经过改进后的伴随 DAG

## 6. 结论与讨论

### 6.1. 结论

1) 通过理论推导和分析从整体方法的角度,展示了伴随方程在变分资料同化中的重要价值,证明了在分析变量维数特别巨大的情况下伴随敏感度计算方法在计算代价方面相对直接计算方法具有明显优势。

2) 结合数值预报模式中所采用的时间/空间离散格式分析了伴随方程中 Jacobian 矩阵的稀疏性及非零元素分布特征:如果采用局部化时间离散格式构造数值预报模式,那么变分资料同化相应伴随方程中的 Jacobian 矩阵将是块对角阵且具有固定块带宽值;采用局部化空间差分格式后,使得 Jacobian 矩阵中的非零矩阵块进一步呈现稀疏特征,甚至呈现出三对角结构;块上三角结构可以实现离散伴随方程的逆向逐步求解策略。

3) 针对变分资料同化中 Jacobian 矩阵的稀疏性,提出了一种压缩存储策略。通过适当调整非零元素的分布位置,可以实现稀疏 Jacobian 矩阵的压缩存储,从而减小存储空间需求。因为在数值预报模式中所采用的时间和空间离散格式固定后,其相应伴随方程的 Jacobian 矩阵稀疏结构不再变化,所以可以采用固定不变的压缩存储策略。

4) 利用有向无环图(ADG)对函数计算过程进行了表示,在此基础上分别利用线性化和伴随 ADG 表示了切线性和伴随计算过程,通过示例说明通过消除伴随计算图中的顶点能够减小计算代价,减少的初等函数计算次数等于消除的顶点个数;采用顶点消除方法对四维变分资料同化极小化过程中调用的热点伴随程序进行优化改进后,将大大减小计算量,从而为伴随方程求解优化奠定基础,最终加速资料同化系统的计算过程。

### 6.2. 讨论

在增量变分资料同化框架中目标泛函及其梯度向量的计算是非常重要的组成部分,两者实质上都是敏感度计算过程。尤其是伴随敏感度计算在时间上是逆向的,存储和计算代价都非常巨大,因此需要研究伴随敏感度计算优化技术。本文结合大气预报方程组离散格式,分析了伴随敏感度计算中 Jacobian 矩阵的稀疏性及非零元素分布特征,从而为伴随敏感度计算给出了优化的技术路径。但在实际业务同化系统设计和实现过程中,还需要针对具体问题具体分析,在实际中伴随敏感度计算是一个非常复杂的问题。

## 基金项目

国家重点研发计划(批准号:2018YFC1506704)和国家自然科学基金(41475094, 41105063, 41375105)资助的课题。

## 参考文献

- [1] 黄思训, 伍荣生. 大气科学中的数学物理问题[M]. 北京: 气象出版社, 2001: 460.
- [2] 邹晓蕾. 资料同化理论与应用(上册)[M]. 北京: 气象出版社, 2009.
- [3] 曹小群, 黄思训, 杜华栋. 变分同化中水平误差函数的正交小波模拟新方法[J]. 物理学报, 2008, 57(3): 1984-1989.
- [4] 韩月琪, 张耀存, 王云峰. 一种新的顺序数据同化方法[J]. 中国科学 E 辑: 技术科学, 2009, 39(8): 1472-1482.
- [5] 曹小群, 宋君强, 张卫民. 一种基于复数域微分的资料同化新方法[J]. 物理学报, 2013, 62(17): 170504-9.
- [6] Evensen, G. (1994) Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, **99**, 10143-10162. <https://doi.org/10.1029/94JC00572>
- [7] Houtekamer, P.L., Mitchell, H.L., Pellerin, G., et al. (2005) Atmospheric Data Assimilation with an Ensemble Kalman

- Filter: Results with Real Observations. *Monthly Weather Review*, **133**, 604-620. <https://doi.org/10.1175/MWR-2864.1>
- [8] 闵锦忠, 刘盛玉, 毕坤, 等. 基于 Hybrid EnSRF-En3DVar 的雷达资料同化研究[J]. 大气科学学报, 2015, 38(2): 213-221.
- [9] 闵锦忠, 尤悦, 高士博, 等. 微物理方案及其参数对 EnSRF 资料同化的影响研究[J]. 大气科学学报, 2016, 39(4): 510-524.
- [10] Rabier, F., Jarvinen, H., Klinker, E., *et al.* (2000) The ECMWF Operational Implementation of Four Dimensional Variational Assimilation. Part I: Experimental Results with Simplified Physics. *Quarterly Journal of the Royal Meteorological Society*, **126**, 1143-1170. <https://doi.org/10.1002/qj.49712656415>
- [11] Courtier, P., Thepaut, J.N. and Hollingsworth, A. (1994) A Strategy for Operational Implementation of 4D-Var, Using an Incremental Approach. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367-1387. <https://doi.org/10.1002/qj.49712051912>
- [12] 张卫民, 曹小群, 宋君强. 以全球谱模式为约束的四维变分资料同化系统 YH4DVAR 的设计和实现[J]. 物理学报, 2012, 61(24): 249202.
- [13] 马旭林, 庄照荣, 薛纪善, 等. GRAPES 非静力数值预报模式的三维变分资料同化系统的发展[J]. 气象学报, 2009, 67(1): 50-60.
- [14] 马旭林, 和杰, 周勃昉, 等. 非高斯分布观测误差资料的变分质量控制对暴雨预报的影响[J]. 大气科学学报, 2017, 40(2): 170-180.
- [15] Giering, R. and Kaminski, T. (1998) Recipes for Adjoint Code Construction. *ACM Transactions on Mathematical Software*, **24**, 437-474. <https://doi.org/10.1145/293686.293695>
- [16] 程强, 曹建文, 王斌, 等. 伴随模式生成器[J]. 中国科学 F 辑: 信息科学, 2009, 39(5): 545-558.
- [17] 刘永柱, 张林, 金之雁. GRAPES 全球切线性和伴随模式的调优[J]. 应用气象学报, 2017, 28(1): 62-71.
- [18] 郑琴, 叶飞辉, 沙建新, 等. 几种约束优化算法求解含“开关”过程的条件非线性最优扰动的比较[J]. 大气科学学报, 2017, 40(2): 273-279.