

真比例导引律的降维分析方法^{*}

黎克波,陈磊,张翼

(国防科技大学 航天与材料工程学院,湖南 长沙 410073)

摘要:提出了三维真比例导引律的降维分析方法。首先研究视线的运动规律,提出了视线的两种旋转角速度及其计算方法,建立了视线运动方程与新的弹目相对运动方程。其次通过对弹目相对运动的分析,发现在三维空间中存在视线瞬时旋转平面,可以在该平面内构造二维制导律,以应用于三维拦截问题的制导控制。然后将空间真比例导引律不加任何近似与线性化直接引入视线瞬时旋转平面,使其维度从三维降为二维,并对其制导特性进行了分析。最后通过仿真对比分析,验证了真比例导引律降维分析方法的有效性。

关键词:导弹制导;视线运动方程;真比例导引;降维分析方法

中图分类号: V448.133 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0001-05

Dimension-reduction method of true proportional navigation guidance law

LI Kebo, CHEN Lei, ZHANG Yi

(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The dimension-reduction method of the true proportional navigation (TPN) guidance law is proposed. Firstly, the rotation principle of the line of sight (LOS) was investigated, two kinds of angular velocities of LOS were presented, and so were their calculation methods. Then, the LOS kinematic equation was advanced, and the new relative dynamic equation between interceptor and target was proposed. It was found that there is an instantaneous rotation plane of LOS (IRPL) in the three-dimensional (3D) space, in which two-dimensional (2D) guidance laws can be constructed to serve for 3D interception guidance. The spatial 3D true proportional navigation (TPN) guidance law was directly introduced into IRPL without any approximation and linearization to transform into dimension-reduced 2D TPN, and the characteristics were studied. Finally, an exoatmospheric interception was taken for simulation to demonstrate the effectiveness of the theory presented in this paper.

Key words: missile guidance; kinematic equation of LOS; true proportional navigation; dimension-reduction method

比例导引律(PN)是目前最为实用和研究最为广泛的制导律,经过几十年的发展,已经得到了PN的多种形式。20世纪90年代以前学者对PN的研究主要集中在二维平面,集大成者是C D Yang和C C Yang^[1],他们提出了比例导引律的一种统一框架,并将其对制导律的数学分析方法推广至三维真比例导引律(TPN)^[2-3]。

20世纪90年代中后期,一些学者开始对三维弹目相对运动与制导律的构造提出新的思想。Duflos等^[4]基于导弹制导的捕获要求,提出了一种制导建模方法,并推导了三维空间中的扩展真比例导引律(GTPN),使其具有与PPN相同的捕获区域。Chiou和Kuo^[5-7]基于古典微分几何曲线理论,将导弹和目标的运动轨迹视为三维空间中的曲线,提出了弧长体系下导弹制导的微分几何制导指令。李超勇等^[8]将Chiou和Kuo所提出

的微分几何制导指令从弧长体系转化至时域内,并推导了在拦截高速目标时的初始捕获必要条件与奇异条件。Tyan^[9]基于修正极坐标系(MPC),利用变量替换构建弹目相对运动方程,提出了新的制导律统一框架,并对TPN、GTPN、扩展理想比例导引律(GIPN)以及PPN的性能和捕获区域进行了分析。以上学者研究的着眼点各不相同,包括:导弹的捕获要求、弧长体系下的微分几何曲线理论、修正极坐标系等。

与以往的文献不同,本文首先对视线的旋转规律进行研究,提出了视线的两种旋转方式和旋转角速度,并建立了视线运动方程。同时结合弹目相对运动,构建了新的相对运动方程。通过对该方程的分析,发现弹目相对运动可以解耦为视线瞬时旋转平面内的相对运动与该平面的转动,并且在视线瞬时旋转平面内构造二维制导律,可

^{*} 收稿日期:2011-10-15

基金项目:国家863计划资助项目

作者简介:黎克波(1986—),男,贵州金沙人,博士研究生,E-mail: likeboreal@qq.com;

陈磊(通信作者),男,教授,博士,硕士生导师,E-mail: clwhl@263.net

以解决三维拦截问题。本文将三维 TPN 不加任何近似与线性化,直接引入视线瞬时旋转平面,成为降维 TPN,并通过拦截高速大机动目标的仿真算例,验证真比例导引律降维分析方法的有效性。

1 视线运动方程

拦截弹和目标在发射惯性系中运动,其相对位置关系如图 1 所示。

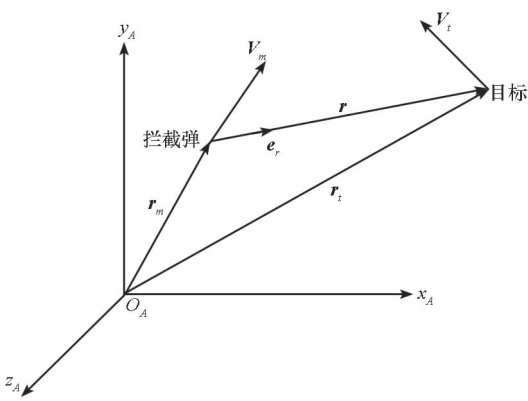


图 1 弹目相对位置关系

Fig. 1 The relative position between interceptor and target
由图可知

$$\mathbf{r} = \mathbf{r}_t - \mathbf{r}_m = r\mathbf{e}_r \quad (1)$$

其中, r 为弹目相对距离, \mathbf{e}_r 为弹目相对视线方向单位矢量。 \mathbf{e}_r 在空间中的旋转是三维的, 但若不考虑其绕自身的旋转, 则降低为二维。

设视线瞬时旋转角速度矢量为 $\boldsymbol{\omega}_s = \omega_s \mathbf{e}_\omega$, \mathbf{e}_ω 为该角速度的方向, 有

$$\frac{d\mathbf{e}_r}{dt} = \boldsymbol{\omega}_s \times \mathbf{e}_r = \omega_s \mathbf{e}_\omega \times \mathbf{e}_r = \omega_s \mathbf{e}_\theta \quad (2)$$

其中, ω_s 即为视线的瞬时转率, 根据定义, $\omega_s \geq 0$ 。 \mathbf{e}_r 、 \mathbf{e}_θ 、 \mathbf{e}_ω 为相互垂直正交的单位向量, 以拦截弹质心为原点, 组成视线旋转坐标系。 \mathbf{e}_r 和 \mathbf{e}_θ 组成视线瞬时旋转平面 (IRPL, Instantaneous Rotation Plane of LOS), \mathbf{e}_ω 为该平面的法向。由(2)式, 有

$$\mathbf{e}_\theta = \mathbf{e}_\omega \times \mathbf{e}_r \quad (3)$$

在三维空间中, 视线瞬时旋转平面会发生转动, 因而该平面的法向 \mathbf{e}_ω 也会发生旋转。由于 \mathbf{e}_ω 是 \mathbf{e}_r 角速度方向的单位矢量, 因此只能在垂直于 \mathbf{e}_r 的平面内发生旋转。设 \mathbf{e}_ω 的旋转角速度为 $\boldsymbol{\Omega}_s = \Omega_s \mathbf{e}_r$, 则有

$$\frac{d\mathbf{e}_\omega}{dt} = \boldsymbol{\Omega}_s \times \mathbf{e}_\omega = \Omega_s \mathbf{e}_r \times \mathbf{e}_\omega = -\Omega_s \mathbf{e}_\theta \quad (4)$$

根据定义, Ω_s 可正可负。

对上式求导, 有

$$\frac{d\mathbf{e}_\theta}{dt} = \frac{d}{dt}(\mathbf{e}_\omega \times \mathbf{e}_r)$$

$$= \frac{d\mathbf{e}_\omega}{dt} \times \mathbf{e}_r + \mathbf{e}_\omega \times \frac{d\mathbf{e}_r}{dt} \quad (5)$$

将(2)式和(4)式代入(5)式, 有

$$\frac{d\mathbf{e}_\theta}{dt} = -\omega_s \mathbf{e}_r + \Omega_s \mathbf{e}_\omega \quad (6)$$

将(2)、(4)式和(6)式联合起来, 可以得到视线的运动方程:

$$\begin{cases} \frac{d\mathbf{e}_r}{dt} = \omega_s \mathbf{e}_\theta \\ \frac{d\mathbf{e}_\theta}{dt} = -\omega_s \mathbf{e}_r + \Omega_s \mathbf{e}_\omega \\ \frac{d\mathbf{e}_\omega}{dt} = -\Omega_s \mathbf{e}_\theta \end{cases} \quad (7)$$

由以上分析可知, 视线的旋转可以分为两种方式, 即视线在其瞬时旋转平面内的旋转, 角速度大小为 ω_s , 旋转轴为 \mathbf{e}_ω ; 视线瞬时旋转平面的旋转, 角速度大小为 Ω_s , 旋转轴为 \mathbf{e}_r 。 ω_s 即为视线的瞬时转率, Ω_s 即为视线瞬时旋转平面的瞬时转率。

2 弹目相对运动方程

对(1)式求导, 有

$$\dot{r}\mathbf{e}_r + r\frac{d\mathbf{e}_r}{dt} = \dot{\mathbf{r}} = \mathbf{V} = \mathbf{V}_t - \mathbf{V}_m \quad (8)$$

其中, \mathbf{V}_t 为目标速度矢量, \mathbf{V}_m 为拦截弹速度矢量, \mathbf{V} 为弹目相对速度矢量。将(2)式和(7)式代入(8)式, 有

$$\dot{r}\mathbf{e}_r + r\omega_s \mathbf{e}_\theta = \mathbf{V}_t - \mathbf{V}_m = \mathbf{V} \quad (9)$$

由(1)式和(9)式可知 \mathbf{e}_r 、 \mathbf{e}_θ 的计算方法:

$$\mathbf{e}_r = \frac{\mathbf{r}}{|\mathbf{r}|}, \mathbf{e}_\theta = \frac{\mathbf{V} - (\mathbf{V} \cdot \mathbf{e}_r)\mathbf{e}_r}{|\mathbf{V} - (\mathbf{V} \cdot \mathbf{e}_r)\mathbf{e}_r|} \quad (10)$$

对(9)式继续求导, 有

$$\mathbf{a}_t - \mathbf{a}_m = \ddot{r}\mathbf{e}_r + \dot{r}\frac{d\mathbf{e}_r}{dt} + (\dot{r}\omega_s + r\dot{\omega}_s)\mathbf{e}_\theta + r\omega_s \frac{d\mathbf{e}_\theta}{dt} \quad (11)$$

其中, \mathbf{a}_t 为目标的加速度, \mathbf{a}_m 为拦截弹的加速度。将视线的运动方程(7)代入(11)式, 有

$$\mathbf{a}_t - \mathbf{a}_m = (\ddot{r} - r\omega_s^2)\mathbf{e}_r + (r\dot{\omega}_s + 2\dot{r}\omega_s)\mathbf{e}_\theta + r\omega_s \Omega_s \mathbf{e}_\omega \quad (12)$$

(12)式即为视线旋转坐标系内的弹目相对运动方程, 可以改写为如下标量方程组:

$$\begin{cases} \ddot{r} - r\omega_s^2 = a_{tr} - a_{mr} \\ r\dot{\omega}_s + 2\dot{r}\omega_s = a_{t\theta} - a_{m\theta} \\ r\omega_s \Omega_s = a_{t\omega} - a_{m\omega} \end{cases} \quad (13)$$

下标“ r, θ, ω ”分别表示沿视线旋转坐标系三轴的分量。该相对运动方程与经典的视线坐标系内的

方程^[11]相比,形式更加简洁,且没有三角函数项。

3 视线两种旋转角速度的计算

视线旋转坐标系 $\{e_r, e_\theta, e_\omega\}$ 与视线坐标系 $o_A - x_s y_s z_s$ 的关系如图2所示。

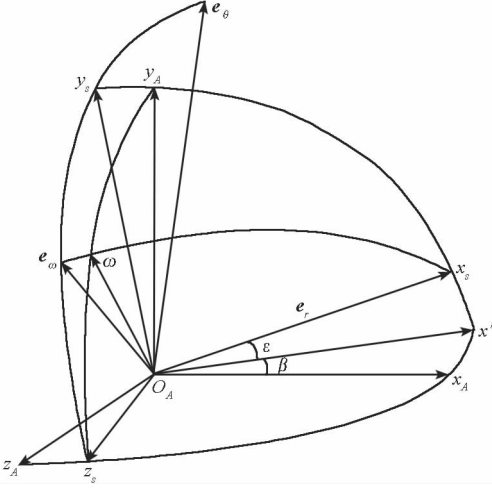


图2 视线旋转坐标系、视线坐标系与发射惯性系之间的关系

Fig. 2 The relationship among the LOS rotation coordinates, the LOS coordinates and the launch inertial coordinates

图2中,发射惯性系 $o_A - x_A y_A z_A$ 通过“23”旋转得到视线坐标系 $o_A - x_s y_s z_s$,欧拉角分别为:视线高低角 ε 、视线方位角 β 。于是有

$$\begin{cases} x_s = \cos\varepsilon\cos\beta \cdot x_A + \sin\varepsilon \cdot y_A \\ \quad - \cos\varepsilon\sin\beta \cdot z_A \\ y_s = -\sin\varepsilon\cos\beta \cdot x_A + \cos\varepsilon \cdot y_A \\ \quad + \sin\varepsilon\sin\beta \cdot z_A \\ z_s = \sin\beta \cdot x_A + \cos\beta \cdot z_A \end{cases} \quad (14)$$

其中, x_A, y_A, z_A 表示发射惯性系三轴的方向, x_s, y_s, z_s 表示视线坐标系三轴的方向。

设视线坐标系的旋转角速度 ω 为

$$\begin{aligned} \omega &= \dot{\beta} + \dot{\varepsilon} = \dot{\beta}y_A + \dot{\varepsilon}z_s \\ &= \dot{\beta}\sin\varepsilon \cdot x_s + \dot{\beta}\cos\varepsilon \cdot y_s + \dot{\varepsilon}z_s \end{aligned} \quad (15)$$

虽然 ω 在 x_s 方向上有分量,但由于不需要考虑视线绕自身的滚动旋转,因此视线的旋转角速度(除去绕 x_s 轴的滚动角速度)可为

$$\omega_s = \dot{\beta}\cos\varepsilon \cdot y_s + \dot{\varepsilon}z_s \quad (16)$$

由(2)式可知

$$\begin{cases} \omega_s = \sqrt{(\dot{\beta}\cos\varepsilon)^2 + \dot{\varepsilon}^2} \\ e_\omega = \frac{\omega_s}{\omega_s} = \frac{\dot{\beta}\cos\varepsilon \cdot y_s + \dot{\varepsilon}z_s}{\sqrt{(\dot{\beta}\cos\varepsilon)^2 + \dot{\varepsilon}^2}} \end{cases} \quad (17)$$

于是有

$$e_\theta = e_\omega \times e_r = \frac{\dot{\varepsilon}y_s - \dot{\beta}\cos\varepsilon \cdot z_s}{\sqrt{(\dot{\beta}\cos\varepsilon)^2 + \dot{\varepsilon}^2}} \quad (18)$$

对(16)式求导,有

$$\begin{aligned} \dot{\omega}_s &= \dot{\beta}\cos\varepsilon \cdot y_s - \dot{\beta}\dot{\varepsilon}\sin\varepsilon \cdot y_s \\ &\quad + \dot{\beta}\cos\varepsilon \cdot \dot{y}_s + \ddot{\varepsilon}z_s + \dot{\varepsilon}\dot{z}_s \\ &= (\dot{\beta}\cos\varepsilon - 2\dot{\beta}\dot{\varepsilon}\sin\varepsilon) \cdot y_s \\ &\quad + (\dot{\beta}^2\sin\varepsilon\cos\varepsilon + \ddot{\varepsilon}) \cdot z_s \end{aligned} \quad (19)$$

再对 ω_s 按视线的运动方式求导,有

$$\begin{aligned} \dot{\omega}_s &= \frac{d}{dt}(\omega_s e_\omega) = \dot{\omega}_s e_\omega + \omega_s \frac{de_\omega}{dt} \\ &= \dot{\omega}_s e_\omega + \omega_s \Omega_s \times e_\omega \\ &= \dot{\omega}_s e_\omega - \omega_s \Omega_s e_\theta \end{aligned} \quad (20)$$

可得视线转率导数与视线挠率的计算表达式:

$$\begin{cases} \dot{\omega}_s = \dot{\omega}_s \cdot e_\omega = \frac{\dot{\beta}\dot{\beta}\cos^2\varepsilon - \dot{\beta}^2\dot{\varepsilon}\sin\varepsilon\cos\varepsilon + \ddot{\varepsilon}\dot{\varepsilon}}{\sqrt{(\dot{\beta}\cos\varepsilon)^2 + \dot{\varepsilon}^2}} \\ \Omega_s = \frac{-\dot{\omega}_s \cdot e_\theta}{\omega_s} \\ \quad = \frac{(2\dot{\varepsilon}^2 - \dot{\beta}^2\cos^2\varepsilon)\dot{\beta}\sin\varepsilon - (\dot{\beta}\dot{\varepsilon} + \ddot{\varepsilon}\dot{\beta})\cos\varepsilon}{\sqrt{(\dot{\beta}\cos\varepsilon)^2 + \dot{\varepsilon}^2}} \end{cases} \quad (21)$$

根据以上分析, $e_r, e_\theta, e_\omega, \omega_s, \dot{\omega}_s, \Omega_s$ 皆可以表示为 $\varepsilon, \beta, \dot{\varepsilon}, \dot{\beta}, \ddot{\varepsilon}, \ddot{\beta}$ 的函数。导引头可以通过测量视线高低角 ε 和视线方位角 β ,并进行相应滤波,计算视线运动方程的各参数。

4 真比例导引律的降维

本文采用 Duflos^[4]对目标捕获的定义:拦截弹对目标的捕获是拦截弹通过制导控制,在有限时间 t_f 内,使弹目相对距离 $r(t_f)$ 降低至0。表达式如下:

$$\exists t_f < \infty, \quad r(t_f) = 0 \quad (22)$$

为了避免视线转率趋于无穷大,还需满足如下条件:

$$r(t_f)\omega_s(t_f) = 0 \quad (23)$$

由相对运动方程(13)式可知,方程组:

$$\begin{cases} \ddot{r} - r\omega_s^2 = a_{tr} - a_{mr} \\ r\dot{\omega}_s + 2\dot{r}\omega_s = a_{t\theta} - a_{m\theta} \end{cases} \quad (24)$$

决定弹目相对距离 r 和瞬时视线转率 ω_s 的变化,而方程

$$r\omega_s\Omega_s = a_{t\omega} - a_{m\omega} \quad (25)$$

决定视线瞬时旋转平面的旋转角速率 Ω_s 的变化。 r, ω_s 与 Ω_s 的计算是解耦的。于是,通过该方程可以将弹目相对运动的二阶非线性耦合常微分方程组解耦为两种运动:视线瞬时旋转平面内的相对运动;视线瞬时旋转平面的转动。

根据捕获的定义,即使在拦截末段 $\Omega_s \neq 0$,只

要可以控制视线转率 ω_s 收敛,则导弹仍然可以捕获目标。并且,视线旋转坐标系 $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\omega$ 的旋转总是趋向于减小相对加速度在 \mathbf{e}_ω 上的分量,使 Ω_s 随弹目相对距离的减小而趋于某一较小的有界值。由于相对运动方程(13)高度非线性,将通过数值仿真验证这一理论。根据以上分析,可以忽略拦截弹在 \mathbf{e}_ω 方向上的控制而使 $a_{m\omega} = 0$ 。

由(24)式的第一式,有

$$\ddot{r} = a_{tr} - a_{mr} + r\omega_s^2 \quad (26)$$

对(26)式求导:

$$\dot{r} = \frac{d(a_{tr} - a_{mr})}{dt} + 2\omega_s(a_{t\theta} - a_{m\theta}) - 3\dot{r}\omega_s^2 \quad (27)$$

根据(26)、(27)两式可知,当存在不可忽略的 ω_s 时,弹目相对加速度沿 \mathbf{e}_r 方向的分量“ $a_{tr} - a_{mr}$ ”对 r 的影响与沿 \mathbf{e}_θ 方向上的分量“ $a_{t\theta} - a_{m\theta}$ ”相比,是一个高阶小量;当 ω_s 趋近于 0 时,由于弹目相对速度 \dot{r} 足够小,可以忽略“ $a_{tr} - a_{mr}$ ”对拦截产生的影响。因此对于以最终捕获目标为目的的导弹,可以不控制沿视线方向的运动,使 $a_{mr} = 0$ 。于是,根据对弹目相对运动方程的分析,解决制导问题的关键是求解 \mathbf{e}_θ 方向上的制导指令加速度 $a_{m\theta}$ 。

传统的空间制导律通常选择在视线坐标系内建立三维弹目相对运动方程^[11]:

$$\begin{cases} \ddot{r} - r(\omega_{ys}^2 + \omega_{zs}^2) = a_{tx} - a_{mx} \\ r\dot{\omega}_{zs} + 2\dot{r}\omega_{zs} + r\omega_{xs}\omega_{ys} = a_{ty} - a_{my} \\ r\dot{\omega}_{ys} + 2\dot{r}\omega_{ys} - r\omega_{xs}\omega_{zs} = -a_{tz} + a_{mz} \end{cases} \quad (28)$$

其中, $\omega_{xs} = \beta \sin \varepsilon$, 为 ω 在 x_s 轴的分量; $\omega_{ys} = \beta \cos \varepsilon$, 为 ω 在 y_s 轴的分量; $\omega_{zs} = \dot{\varepsilon}$, 为 ω 在 z_s 轴的分量; $[a_{tx}, a_{ty}, a_{tz}]^T, [a_{mx}, a_{my}, a_{mz}]^T$ 为目标与拦截弹的加速度在视线坐标系内的表示。空间 TPN 制导指令表达式如下:

$$\begin{cases} a_{mx} = 0 \\ a_{my} = N|\dot{r}|\omega_{zs} \\ a_{mz} = -N|\dot{r}|\omega_{ys} \end{cases} \quad (29)$$

其中, N 为有效导航比,满足比例导引弹道收敛条件: $N > 2$ 。TPN 在 y_s, z_s 上的两个制导指令加速度是相互独立的,其共同作用的结果相当于在拦截弹上作用了一个合加速度:

$$\begin{aligned} \mathbf{a}_m &= -N\dot{r}\dot{\varepsilon} \cdot \mathbf{y}_s + N\dot{r}\beta \cos \varepsilon \cdot \mathbf{z}_s \\ &= N\dot{r} \sqrt{(\beta \cos \varepsilon)^2 + \dot{\varepsilon}^2} \cdot \left(\frac{-\dot{\varepsilon} \cdot \mathbf{y}_s + \beta \cos \varepsilon \cdot \mathbf{z}_s}{\sqrt{(\beta \cos \varepsilon)^2 + \dot{\varepsilon}^2}} \right) \end{aligned} \quad (30)$$

将(17)式和(18)式代入(30)式,有

$$\mathbf{a}_m = -N\dot{r}\omega_s \cdot \mathbf{e}_\theta \quad (31)$$

(31)式中, ω_s 为视线在其瞬时旋转平面内的转率, \mathbf{e}_θ 为视线法向,同样位于视线瞬时旋转平面内。因此,三维 TPN 可以不加任何近似与线性化,直接转化成视线瞬时旋转平面内的二维 TPN。由(17)式和(18)式可知,真比例导引律的降维只需要视线高低角和方位角的一阶微分信息,可以应用于工程实际。

将(31)式代入(24)式的第二式可得

$$r\dot{\omega}_s + 2\dot{r}\omega_s = a_{t\theta} + N\dot{r}\omega_s \quad (32)$$

若初始时刻视线转率为 ω_{s0} ,且目标不机动,可以由(32)式推导出

$$r\dot{\omega}_s = (N-2)\dot{r}\omega_s \quad (33)$$

进一步推导出

$$\omega_s = \omega_{s0} \left(\frac{r}{r_0} \right)^{(N-2)} \quad (34)$$

由(34)式可知,当 $N > 2$ 时,对非机动目标,TPN 可以使视线瞬时旋转平面内的视线转率随弹目相对距离的减小而逐渐降低至 0。这与平面真比例导引律的性能分析相一致。至此,空间三维 TPN 已经转化为视线瞬时旋转平面内的二维 TPN,可以直接采用对平面 TPN 的分析方法,来研究三维 TPN 的性能。根据 Guelman 对平面 TPN 的研究^[12],TPN 对非机动目标的初始捕获条件为

$$(r_0\omega_{s0})^2 \leq (2N-1)r_0^2 \quad (35)$$

(35)式对于视线瞬时旋转平面内的降维 TPN 而言仍然成立。三维真比例导引律通过降维以后,可以应用对平面 TPN 的分析方法对其性能进行研究,降低了分析问题的难度。

5 仿真分析

以拦截大气层外高速机动目标仿真为例,初始状态参数如表 1 所示。

表 1 拦截初始条件

Tab. 1 Inertial interception condition			
拦截弹	X	Y	Z
位置(km)	2837.575	735.590	119.855
速度(km/s)	5.055	-0.872	0.158
目标	X	Y	Z
位置(km)	2882.410	735.033	80.905
速度(km/s)	1.156	0.337	5.621

设目标在发射惯性系的 $\mathbf{y}_A, \mathbf{z}_A$ 方向有常加速度 $\mathbf{a}_t = [0 \quad 4g_0 \quad 4g_0]^T$,制导系数 $N = 4$,不考虑视线角测量噪声,取导引头采样周期为 10ms,盲区为 500m,计算脱靶量。仿真结果如图 3 和图 4 所示。

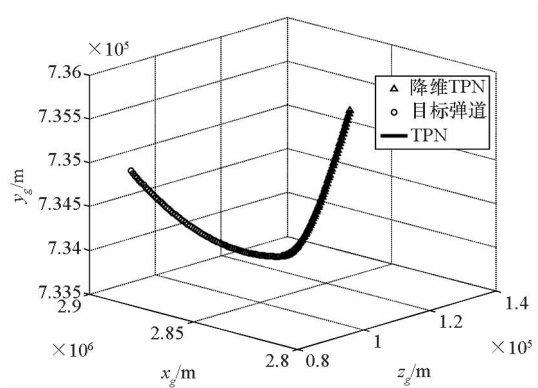


图3 空间TPN与降维TPN的三维拦截弹道
Fig.3 The 3-dimensional interception trajectories of 3-dimensional TPN and dimension-degenerated TPN

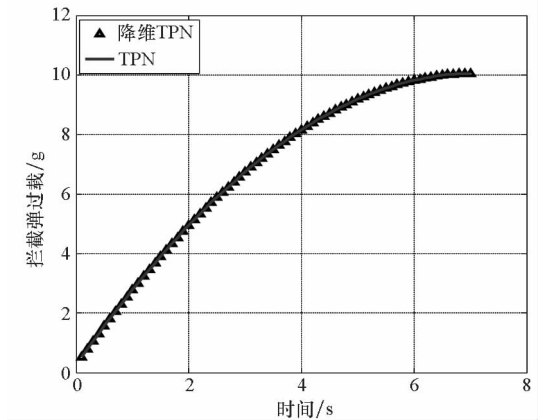


图4 空间TPN与降维TPN的瞬时过载
Fig.4 The instantaneous overloads of TPN and dimension-degenerated TPN

从以上仿真结果可以发现,三维TPN和降维TPN具有几乎一致的拦截弹道和过载分布。仿真结果显示,两种TPN的脱靶量皆为0.1760m。

视线瞬时旋转平面的旋转角速度 Ω_z 在拦截过程中的变化趋势如图5所示。

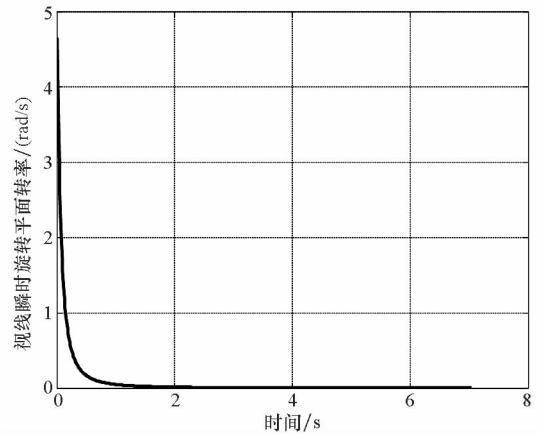


图5 IRPL瞬时转率的变化
Fig.5 The instantaneous IRPL rate of TPN

由图可知,IRPL的瞬时转率在拦截过程中逐渐降低至0,代表了视线瞬时旋转平面逐渐趋于

稳定的过程。

6 结 论

本文基于对视线旋转基本原理的研究,提出了视线的两种旋转角速度及其计算方法,得到了视线运动方程。将视线运动与弹目相对运动结合起来,得到了新的弹目相对运动方程。通过对弹目相对运动方程的分析,发现存在视线瞬时旋转平面,在该平面内可以构造二维制导律,应用于三维拦截问题。

将空间真比例导引律不加近似与线性化地直接引入视线瞬时旋转平面,得到了具有二维形式的降维TPN,可以应用研究平面TPN的经典方法对其性能进行分析,降低了分析问题的难度。通过数值仿真,验证了真比例导引律降维分析方法的有效性。

参考文献(References)

[1] Yang C D, Yang C C. A unified approach to proportional navigation[J]. IEEE Transactions on Aerospace and Electronic Systems, 1997, 33(2):557-567.

[2] Yang C D, Yang C C. Analytical solution of 3D true proportional navigation[J]. IEEE Transactions on Aerospace and Electronic Systems, 1996, 32(4): 1509-1522.

[3] Yang C D, Yang C C. Analytical solution of three-dimensional realistic true proportional navigation[J]. Journal of Guidance, Control and Dynamics. 1996,19(3):569-577.

[4] Duflos E, Penel P, Vanheeghe P. 3D guidance law modeling[J]. IEEE Transactions on Aerospace and Electronic Systems. 1999, 35(1): 72-83.

[5] Chiou Y C, Kuo C Y. Geometric approach to three dimensional missile guidance problems[J]. Journal of Guidance, Control, and Dynamics, 1998, 21(2): 335-341.

[6] Kuo C Y, Chiou Y C. Geometric analysis of missile guidance command [J]. IEEE Proceedings: Control Theory and Applications, 2000, 147(2):205-211.

[7] Kuo C Y, Soetanto D, Chiou Y C. Geometric analysis of flight control command for tactical missile guidance [J]. IEEE Transactions on Control Systems Technology, 2001, 9(2):234-243.

[8] 李超勇,荆武兴,齐治国,等. 空间微分几何制导律应用研究[J]. 宇航学报, 2007, 28(5): 1235-1240.
LI Chaoyong, JING Wuxin, QI Ziguo, et al. Application of the 3D differential geometric guidance commands[J]. Journal of Astronautics, 2007, 28(5): 1235-1240. (in Chinese)

[9] Tyan F. The capture region of a general 3D TPN guidance law for missile and target with limited maneuverability [C]// Proceedings of the American Control Conference, Arlington, VA June 25-27, 2001.

[10] Tyan F. Unified approach to missile guidance laws: a 3D extension [J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(4): 1178-1199.

[11] Chen L, Zhang B. Novel TPN control algorithm for exoatmospheric intercept[J]. Journal of Systems Engineering and Electronics, 2009, 20(6): 1290-1295.

[12] Guelman M. The closed-form solution of true proportional navigation [J]. IEEE Transactions on Aerospace and Electronic Systems, 1976, 12(4): 472-482.

线粘弹性材料中三维裂纹问题的加料有限元法*

段静波,雷勇军

(国防科技大学 航天与材料工程学院,湖南 长沙 410073)

摘要: 将加料有限元法扩展应用于线粘弹性材料三维断裂问题。为了反映裂纹尖端的奇异性,在裂尖附近的奇异区采用若干八节点六面体加料单元和过渡单元,非奇异区采用常规八节点六面体单元。三种单元分区混合使用形成求解域网格划分。基于 Boltzmann 叠加原理,推导了粘弹性材料的增量型本构关系,进而获得了增量加料有限元列式,并通过附加自由度计算粘弹性介质中裂纹应变能量释放率。数值算例验证了方法的正确性和有效性。

关键词: 线粘弹性体;三维断裂问题;六面体加料单元;六面体过渡单元;应变能量释放率

中图分类号: O346.1 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0006-06

The enriched finite element method for 3-D fracture problems in viscoelastic materials

DUAN Jingbo, LEI Yongjun

(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The enriched finite element method was developed for three-dimensional fracture problems in a linear viscoelastic body. To manifest the singularity at the crack tip, the 8-node hexahedral enriched elements and corresponding transition elements were employed, combined with ordinary elements on the zone far away from the crack tip. Three types of elements were used together to form the whole mesh. Based on the boltzmann superposition principle, the incremental constitutive relation for viscoelastic materials was formulated. Furthermore, the incremental formulations of the enriched FEM were derived. The strain energy release rate in a cracked viscoelastic body was obtained through the enriched degree of freedoms. The numerical results show that the present method is accurate and efficient.

Key words: linear viscoelasticity; three-dimensional fracture problems; hexahedral enriched element; hexahedral transition element; strain energy release rate

“Enriched finite element”在文献[1]中译为“加料单元”,作者沿用其译法。关于这种方法的研究,1974年,Benzley^[2]最早提出采用加料单元和过渡单元解决各向同性线弹性介质中的裂纹问题,他的基本思想是将裂纹尖端渐近位移场作为加料项引入常规单元位移模式,进而构成加料单元的位移模式。这种方法最大的优点是应力强度因子可以直接从附加自由度中获得,不需要利用节点位移或单元应力数据后处理来获得裂纹断裂参量,这在处理复杂三维裂纹问题时,优势十分明显。目前,这种方法已广泛应用于解决弹性介质中的裂纹问题。Biggers、Pageau 和 Joseph^[3-8]将该方法成功应用到各向异性材料中反平面裂纹、三维楔形切口等几何或材料不连续结构中。Nied等^[9]还将加料有限元法与罚函数相结合解决基

于接触模型的界面裂纹问题。Ayhan^[10-11]采用三维加料有限元方法求解表面裂纹、角裂纹等复杂的三维裂纹问题。最近,Ayhan^[12-13]又将加料有限元法应用于功能梯度材料中的裂纹问题。

本文在前人研究的基础上,将加料有限元法扩展应用于线粘弹性材料三维断裂问题,推导了三维加料有限元单元计算格式和线粘弹性材料增量本构关系,并完成了程序编制。数值算例验证了本文方法的正确性和有效性。

1 加料单元构造

1.1 三维裂纹尖端渐近位移场

通常粘弹性材料的泊松比可视为常数,在此条件下,根据对应原理可得到线粘弹性材料中平面应变裂纹问题、反平面裂纹问题中裂尖渐近位

* 收稿日期:2011-10-21

基金项目: 教育部新世纪优秀人才支持计划资助项目(NCET-08-0148);高等学校博士学科点专项科研基金资助项目(20069998002);国家自然科学基金资助项目(11132012);国防科技大学科研计划资助项目(JC10-01-01)

作者简介: 段静波(1982—),男,山西晋城人,博士研究生,E-mail:duanjingbo@nudt.edu.cn;

雷勇军(通信作者),男,教授,博士,博士生导师,E-mail:leiyj108@nudt.edu.cn

移场

$$\begin{cases} u' = \frac{1}{4} \sqrt{\frac{2r}{\pi}} \left((5-3v-8v^2) \cos \frac{\theta}{2} - (1+v) \cos \frac{3\theta}{2} \right) L^{-1} [K_I^*/E_{\sim}^*] \\ \quad + \frac{1}{4} \sqrt{\frac{2r}{\pi}} \left((9+v-8v^2) \sin \frac{\theta}{2} + (1+v) \sin \frac{3\theta}{2} \right) L^{-1} [K_{II}^*/E_{\sim}^*] \\ v' = \frac{1}{4} \sqrt{\frac{2r}{\pi}} \left((7-v-8v^2) \sin \frac{\theta}{2} - (1+v) \sin \frac{3\theta}{2} \right) L^{-1} [K_I^*/E_{\sim}^*] \\ \quad + \frac{1}{4} \sqrt{\frac{2r}{\pi}} \left(-(3-5v-8v^2) \cos \frac{\theta}{2} - (1+v) \cos \frac{3\theta}{2} \right) L^{-1} [K_{II}^*/E_{\sim}^*] \\ w' = 2(1+v) \sqrt{\frac{2r}{\pi}} \sin \frac{\theta}{2} L^{-1} [K_{III}^*/E_{\sim}^*] \end{cases} \quad (1)$$

其中, u', v', w' 分别为沿图 1 中所示局部直角坐标系坐标轴 x', y', z' 方向的位移, r, θ, z' 为以裂尖 O' 点为原点的局部柱坐标系坐标轴, $K_I^*, K_{II}^*, K_{III}^*$ 为拉普拉斯变换后的 I、II、III 型应力强度因子, $E_{\sim}^* = pE^*$, E^* 为拉普拉斯变换后的松弛模量, p 为拉普拉斯变换参数, $L^{-1}[\cdot]$ 为拉普拉斯逆变换, v 为泊松比。

$Q'(r, \theta) =$

$$\frac{1}{4} \sqrt{\frac{2r}{\pi}} \begin{bmatrix} (5-3v-8v^2) \cos \frac{\theta}{2} - (1+v) \cos \frac{3\theta}{2} & (9+v-8v^2) \sin \frac{\theta}{2} + (1+v) \sin \frac{3\theta}{2} & 0 \\ (7-v-8v^2) \sin \frac{\theta}{2} - (1+v) \sin \frac{3\theta}{2} & -(3-5v-8v^2) \cos \frac{\theta}{2} - (1+v) \cos \frac{3\theta}{2} & 0 \\ 0 & 0 & 8(1+v) \sin \frac{\theta}{2} \end{bmatrix} \quad (2)$$

那么, 裂纹前缘任意一点 O' 处局部渐近位移场在总体直角坐标系下为

$$\{u \ v \ w\}^T = Q(r, \theta) \{L^{-1}[K_I^*/E_{\sim}^*] \ L^{-1}[K_{II}^*/E_{\sim}^*] \ L^{-1}[K_{III}^*/E_{\sim}^*]\}^T \quad (3)$$

其中, u, v, w 分别为沿图 1 中所示总体直角坐标系坐标轴 x, y, z 方向的位移, $Q(r, \theta) = T^T Q'(r, \theta)$, T 为局部直角坐标系与整体直角坐标系间的转换矩阵。

1.2 单元位移模式

在普通八节点六面体等参单元位移模式基础上, 通过加入表征裂纹尖端奇异特性的(3)式, 可构成加料单元的位移模式

$$u_i = \sum_{m=1}^8 N_m(\xi, \eta, \zeta) \bar{u}_{im} + Z_0(\xi, \eta, \zeta) \sum_{j=1}^3 [Q_{ij}(r, \theta) \psi_j(\Gamma) - \sum_{m=1}^8 N_m \bar{Q}_{ijm}(r, \theta) \bar{\psi}_j(\Gamma)] \quad (4)$$

其中, u_i 为单元内一点在整体坐标系中沿坐标轴

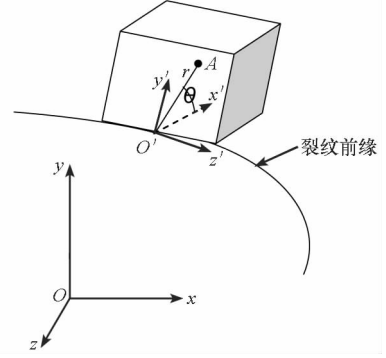


图 1 三维裂纹前缘的裂纹单元

Fig. 1 The crack element on the three-dimensional crack front

对于各向同性线粘弹性材料中一般的三维裂纹问题, 虽然结构几何尺寸、大小、载荷作用方式、裂纹形状与前缘不尽相同, 但在裂纹前缘任意一点 O' 处(图 1 所示), 其局部渐近位移场可由平面应变问题的裂尖渐近位移场与反平面问题的裂尖渐近位移场组合而成, 且令

方向的位移, $i = 1, 2, 3, \bar{u}_{im}$ 为单元节点位移, N_m 为单元形函数, $m = 1, 2, \dots, 8, Q_{ij}(r, \theta)$ 为角函数阵 $Q(r, \theta)$ 的分量, $\bar{Q}_{ijm}(r, \theta)$ 为角函数阵分量在 m 节点处的值, ξ, η, ζ 为等参单元局部坐标系, $\bar{\psi}_j(\Gamma)$ 为 $\psi_j(\Gamma)$ 在 m 节点处的值, $\psi_j(\Gamma)$ 可由插值函数 $N_n(\Gamma)$ 得到

$$\psi_j(\Gamma) = \sum_{n=1}^2 N_n(\Gamma) \psi_j^n \quad (j = 1, 2, 3) \quad (5)$$

其中, Γ 为加料单元位于裂纹前缘线上的边, ψ_j^n 为相应裂纹前缘线上节点的附加自由度, 其值就是未知的 $L^{-1}[K_I^*/E_{\sim}^*]$, $L^{-1}[K_{II}^*/E_{\sim}^*]$, $L^{-1}[K_{III}^*/E_{\sim}^*]$ 。

另外, (3)式中 $Z_0(\xi, \eta, \zeta)$ 为调整函数, 对于裂尖单元而言, $Z_0(\xi, \eta, \zeta) \triangleq 1$, 对于过渡单元, $Z_0(\xi, \eta, \zeta)$ 的表达式取决于过渡单元与裂尖单元相邻的单元面、单元边或单元节点, 例如, 当过渡单元的面 ($\eta = -1$) 与裂尖单元相连时

$$Z_0(\xi, \eta, \zeta) = 0.5(1 - \eta) \quad (6)$$

为了便于理论推导, 将加料单元的位移模式

写成如下的向量形式

$$\mathbf{u} = [N \quad \mathbf{Z}\mathbf{N}_e] \begin{Bmatrix} \bar{\mathbf{u}} \\ \bar{\boldsymbol{\psi}} \end{Bmatrix} \quad (7)$$

其中, $\bar{\mathbf{u}}$ 是加料单元的节点位移列向量, N 是常规六面体单元形函数阵, $\bar{\boldsymbol{\psi}}$ 是加料单元的附加自由度, $\mathbf{Z}\mathbf{N}_e$ 是加料单元形函数阵附加部分。

1.3 裂尖局部坐标系原点的确定

一般来说,对于三维裂纹问题,虽然裂纹前缘线上任意一点 O' 处渐近位移场已给出,但是单元内各点所对应的裂尖局部坐标系并不处处相同。因此,推导单元刚度矩阵前的一个关键问题就是必须首先建立单元内部任意一点 $A(\xi_A, \eta_A, \zeta_A)$ 所对应的局部坐标系 $O' - x'y'z'$, 也就是确定其对应裂尖局部坐标系原点 O' 在裂纹前缘线上的位置。

根据裂尖渐近位移场特性可知,裂尖前缘曲线在裂尖局部坐标系原点 O' 处的切向量,记为 \mathbf{T} , 必须垂直于向量 \mathbf{AO}' , 即

$$\mathbf{AO}' \cdot \mathbf{T} = 0 \quad (8)$$

利用此几何关系就可确定 O' 点在总体坐标系下的坐标值。

2 线粘弹性增量型本构关系

根据 Boltzmann 叠加原理^[14],粘弹性材料的当前应力张量取决于过去历程,其积分型本构关系可写为

$$\boldsymbol{\sigma}(t) = \mathbf{D} E(0) \boldsymbol{\varepsilon}(t) + \mathbf{D} \int_0^t \boldsymbol{\varepsilon}(t-s) \frac{dE(s)}{ds} ds \quad (9)$$

其中, $\boldsymbol{\sigma}(t)$ 、 $\boldsymbol{\varepsilon}(t)$ 分别为 t 时刻的应力向量、应变向量, \mathbf{D} 为材料矩阵, $E(t)$ 为松弛模量,一般情况下可以表示成 Prony 级数形式

$$E(t) = E_\infty + \sum_{n=1}^N E_n e^{-\frac{t}{\tau_n}} \quad (10)$$

为了解决积分函数只能在全历程积分才能求解的困难,有限元分析中通常采用增量型的本构关系,将式(9)的本构方程在时域中离散。在时间 $[0, t]$ 内划分 M 步,步长为 h ,在 $[t_m, t_{m+1}]$ 时间段内, t_{m+1} 时刻的应力应变可以由 t_m 时刻的应力应变来计算,即

$$\begin{cases} \Delta \boldsymbol{\sigma}(t_{m+1}) = \boldsymbol{\sigma}(t_{m+1}) - \boldsymbol{\sigma}(t_m) \\ \Delta \boldsymbol{\varepsilon}(t_{m+1}) = \boldsymbol{\varepsilon}(t_{m+1}) - \boldsymbol{\varepsilon}(t_m) \end{cases} \quad (11)$$

假设在 $[t_m, t_{m+1}]$ 内,应变是线性增加的,即

$$\boldsymbol{\varepsilon}(t) = \boldsymbol{\varepsilon}(t_m) + \frac{t - t_m}{t_{m+1} - t_m} \Delta \boldsymbol{\varepsilon}(t_{m+1}) \quad (12)$$

于是,联立(9)式、(10)式、(11)式与(12)式,可得增量型本构方程

$$\Delta \boldsymbol{\sigma}(t_{m+1}) = \hat{\mathbf{D}} \Delta \boldsymbol{\varepsilon}(t_{m+1}) - \sum_{n=1}^N \alpha_n(h) \boldsymbol{\sigma}^n(t_m) \quad (13)$$

其中, $\Delta \boldsymbol{\sigma}(t_{m+1})$ 为应力增量, $\Delta \boldsymbol{\varepsilon}(t_{m+1})$ 为应变增量, $\hat{\mathbf{D}} = [\mathbf{E}_\infty + \sum_{n=1}^N \beta_n(h) \mathbf{E}_n] \mathbf{D}$, $\alpha_n(h) = 1 - e^{-\frac{h}{\tau_n}}$, $\beta_n(h) = \alpha_n(h) \frac{\tau_n}{h}$, $\boldsymbol{\sigma}^n(t_m)$ 的递推式为 $\boldsymbol{\sigma}^n(t_m) = \mathbf{E}_n \beta_n(h) \mathbf{D} \Delta \boldsymbol{\varepsilon}(t_m) - (\alpha_n(h) - 1) \boldsymbol{\sigma}^n(t_{m-1})$, 且 $\boldsymbol{\sigma}^n(t_0) = 0$ 。

3 有限元增量列式

在单元域内,增量形式的虚功原理为

$$\int_{\Omega^e} \delta \boldsymbol{\varepsilon}^T \Delta \boldsymbol{\sigma} d\Omega = \int_{\Omega^e} \delta \mathbf{u}^T \Delta \mathbf{f} d\Omega + \int_{\Gamma^e} \delta \mathbf{u}^T \Delta \mathbf{p} d\Gamma \quad (14)$$

其中, $\delta \mathbf{u}$ 为虚位移, $\delta \boldsymbol{\varepsilon}$ 为虚应变, $\Delta \mathbf{f}$ 为体力增量, $\Delta \mathbf{p}$ 为面力增量。

将(7)式、(13)式代入(14)式,可得

$$\begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{u\psi} \\ \mathbf{K}_{\psi u} & \mathbf{K}_{\psi\psi} \end{bmatrix} \begin{Bmatrix} \Delta \bar{\mathbf{u}}(t_{m+1}) \\ \Delta \bar{\boldsymbol{\psi}}(t_{m+1}) \end{Bmatrix} = \begin{Bmatrix} \Delta \mathbf{F}_u(t_{m+1}) \\ \Delta \mathbf{F}_\psi(t_{m+1}) \end{Bmatrix} - \begin{Bmatrix} \mathbf{F}_u^0(t_{m+1}) \\ \mathbf{F}_\psi^0(t_{m+1}) \end{Bmatrix} \quad (15)$$

其中,

$$\mathbf{K}_{uu} = \int_{\Omega^e} \mathbf{B}^T \hat{\mathbf{D}} \mathbf{B} d\Omega, \mathbf{K}_{u\psi} = \int_{\Omega^e} \mathbf{B}^T \hat{\mathbf{D}} \mathbf{Z} \mathbf{B}_e d\Omega$$

$$\mathbf{K}_{\psi u} = \int_{\Omega^e} \mathbf{Z} \mathbf{B}_e^T \hat{\mathbf{D}} \mathbf{B} d\Omega, \mathbf{K}_{\psi\psi} = \int_{\Omega^e} \mathbf{Z} \mathbf{B}_e^T \hat{\mathbf{D}} \mathbf{Z} \mathbf{B}_e d\Omega$$

$$\Delta \mathbf{F}_u(t_{m+1}) = \int_{\Omega^e} \mathbf{N}^T \Delta \mathbf{f}(t_{m+1}) d\Omega + \int_{\Gamma^e} \mathbf{N}^T \Delta \mathbf{p}(t_{m+1}) d\Gamma$$

$$\Delta \mathbf{F}_\psi(t_{m+1}) = \int_{\Omega^e} \mathbf{Z} \mathbf{N}_e^T \Delta \mathbf{f}(t_{m+1}) d\Omega + \int_{\Gamma^e} \mathbf{Z} \mathbf{N}_e^T \Delta \mathbf{p}(t_{m+1}) d\Gamma$$

$$\mathbf{F}_u^0(t_{m+1}) = - \int_{\Omega^e} \mathbf{B}^T \sum_{n=1}^N \alpha_n(h) \boldsymbol{\sigma}^n(t_m) d\Omega$$

$$\mathbf{F}_\psi^0(t_{m+1}) = - \int_{\Omega^e} \mathbf{Z} \mathbf{B}_e^T \sum_{n=1}^N \alpha_n(h) \boldsymbol{\sigma}^n(t_m) d\Omega$$

\mathbf{B} 为常规六面体单元应变矩阵, $\mathbf{Z} \mathbf{B}_e$ 为加料单元应变矩阵附加部分。

4 能量释放率

由文献[14]可知,线粘弹性介质中平面裂纹的能量释放率可由(16)式获得

$$\begin{cases} G_I(t) = (1 - \nu^2) K_I(t) L^{-1} [K_I^* / \tilde{E}^*] \\ G_{II}(t) = (1 - \nu^2) K_{II}(t) L^{-1} [K_{II}^* / \tilde{E}^*] \\ G_{III}(t) = (1 + \nu) K_{III}(t) L^{-1} [K_{III}^* / \tilde{E}^*] \end{cases} \quad (16)$$

其中, $G_I(t)$, $G_{II}(t)$, $G_{III}(t)$ 与 $K_I(t)$, $K_{II}(t)$, $K_{III}(t)$ 分别为I、II、III型能量释放率与应力强度因子。

由前面推导可知,(16)式中的 $L^{-1} [K_I^* / \tilde{E}^*]$, $L^{-1} [K_{II}^* / \tilde{E}^*]$, $L^{-1} [K_{III}^* / \tilde{E}^*]$ 就是加料单元的附加自由度 $\bar{\boldsymbol{\psi}}$, 从而可以从有限元列式中直接获得。然后,再通过逆卷积可求得 $K_I(t)$, $K_{II}(t)$,

$K_{III}(t)$,进而由(16)式获得裂纹能量释放率。

5 算例分析

为了验证方法的正确性和有效性,以无限长边裂纹板条问题为例,自编程序,研究了高斯积分点、单元网格大小、加料单元数对裂纹尖端位移以及裂纹能量释放率结果精度的影响规律,并与解析解^[14]作了对比。设板条长度 $H=20a$,宽度 $W=2a$,裂纹长度 $a=1$ 。粘弹性材料采用如图2所示的三参量线粘弹性固体,且假设各参数为 $E_{\infty}=1.0, E_1=1.0, \eta_1=5.0$,泊松比 $\nu=0.3$ 。设板端承受均布拉应力,且 $\sigma_0=1$,载荷历程如图3所示,加载时间为200。图4为边裂纹板条的有限元模型,在板条有限元模型的前后面上约束节点面外位移以模拟平面应变状态。数值算例中为了便于分析对比,暂不涉及物理量单位。

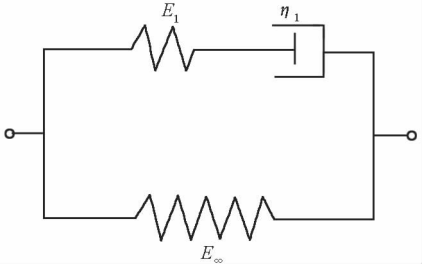


图2 三参量线粘弹性固体
Fig.2 The three-parameter linear viscoelastic solid

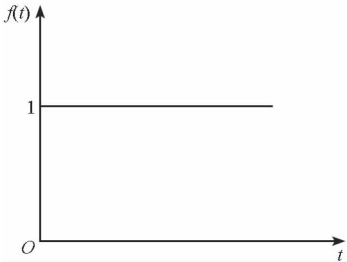
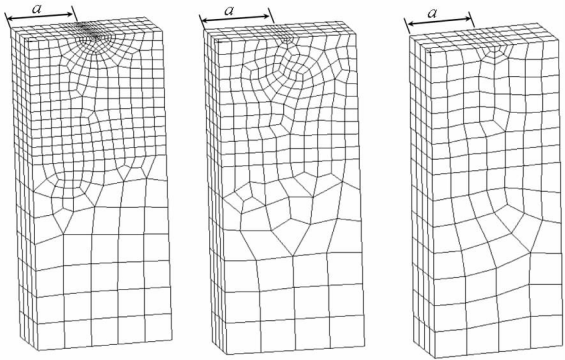


图3 载荷历程
Fig.3 Loading history

(1) 网格密度对位移结果精度的影响

表1中结果主要研究网格密度对位移结果精度的影响。图4(a)~(c)所示的模型分别采用1118单元、680单元、404单元。从表1中可以看出,无论在何种网格密度的情况下,采用本文的加料单元,与常规有限元相比,结果精度都有提高。特别是在较稀疏密度的图4(c)所示的网格下,精度提高幅度较明显。对比3种不同的网格密度下的位移结果,可以发现,加料有限元与常规有限元类似,增大网格密度可以提高位移精度,但是加料有限元精度变化幅度没有常规有限元那么明显,这就是说,加料有限元对网格不是十分敏感,这是其非常好的一个特性。当加料单元边长达到裂纹长度



(a)单元边长0.01a (b)单元边长0.025a (c)单元边长0.05a

图4 不同尺寸裂尖加料单元下的边裂纹板有限元模型
Fig.4 The finite element model for the edge cracked plate with different size of enriched elements

的1/50~1/20时,与解析解相比,误差在5%内,已可满足工程要求,而常规有限元要达到这样的精度,单元边长至少要达到裂纹长度的1/100^[15]。

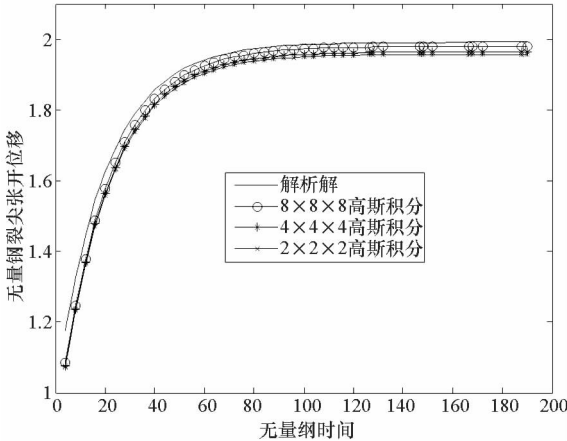


图5 加料单元高斯积分点数对裂尖位移场精度影响
Fig.5 The accuracy comparison for crack opening displacement with different Gauss quadrature points

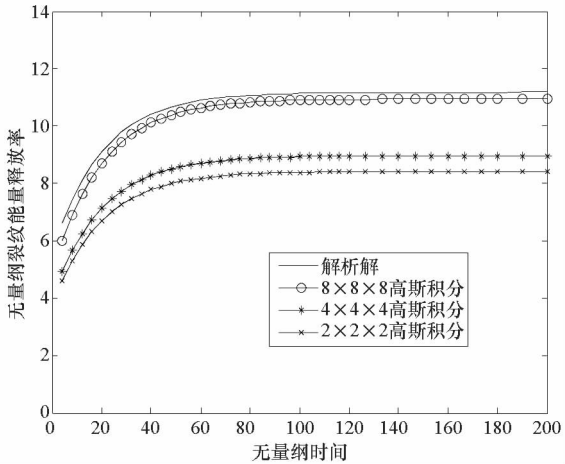


图6 加料单元高斯积分点数对能量释放率精度影响
Fig.6 The accuracy comparison for strain energy release rate with different Gauss quadrature points

表 1 边裂纹板裂尖张开位移 ($r/a=0.15$)
Tab.1 The crack opening displacement for the edge cracked plate ($r/a=0.15$)

网格	无量纲时间	解析解	本文解	误差	误差	误差
网格一 (单元边长 0.01a)	20	1.6248	1.5750	4.98%	1.5612	3.91%
	40	1.8563	1.8289	1.48%	1.8129	2.34%
	80	1.9728	1.9572	0.79%	1.9402	1.65%
	120	1.9885	1.9747	0.69%	1.9575	1.56%
	190	1.9909	1.9766	0.91%	1.9594	1.58%
网格二 (单元边长 0.025a)	20	1.6248	1.5410	5.16%	1.4711	9.46%
	40	1.8563	1.7894	3.60%	1.7083	7.97%
	80	1.9728	1.9150	2.93%	1.8282	7.33%
	120	1.9885	1.9321	2.84%	1.8445	7.24%
	190	1.9909	1.9340	2.86%	1.8463	7.26%
网格三 (单元边长 0.05a)	20	1.6248	1.5008	7.63%	1.3612	16.22%
	40	1.8563	1.7428	6.11%	1.5806	14.85%
	80	1.9728	1.8651	5.46%	1.6915	14.26%
	120	1.9885	1.8817	5.37%	1.7067	14.17%
	190	1.9909	1.8836	5.39%	1.7083	14.19%

表 2 加料单元数对裂尖位移和能量释放率精度的影响
Tab.2 The accuracy comparison for crack opening displacement and strain energy release rate

无量纲时间	裂尖张开位移			能量释放率		
	一层 奇异单元	二层 奇异单元	解析解	一层 奇异单元	二层 奇异单元	解析解
20	1.5157	1.5410	1.6248	7.5292	8.7138	9.3067
40	1.7601	1.7894	1.8563	8.7166	10.1178	10.6328
80	1.8836	1.9150	1.9728	9.3208	10.8308	11.3000
120	1.9005	1.9321	1.9885	9.4034	10.9282	11.3903
190	1.9023	1.9340	1.9909	9.4093	10.9351	11.4040

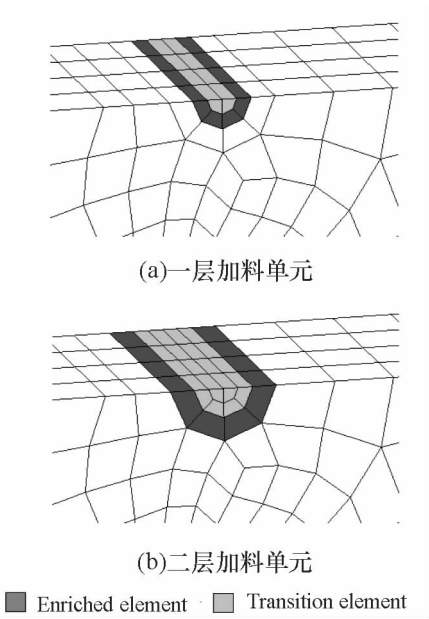


图 7 有限元模型裂尖局部网格
Fig.7 The local mesh for crack tip

(2) 高斯积分点数对位移结果精度的影响

采用图 4(b)所示的模型研究了高斯积分点数对结果影响。图 5 给出的是加料单元采取不同高斯积分点时 $r/a=0.15$ 处裂尖位移结果。由图可知,在加料单元刚度积分采用 $8\times8\times8$ 的减缩积分时获得的结果精度最高,而且加料单元高斯积分点数对位移结果精度影响比较小。图 6 给出的是加料单元采取不同高斯积分点时的能量释放率结果。可以看出,加料单元采取的高斯积分点对其精度有明显的影响,采取 $8\times8\times8$ 的高斯积分点时获得的结果最好,当高斯积分点数减少时,能量释放率结果精度迅速下降。

(3) 加料奇异单元数目对结果精度的影响

仍然采用图 4(b)的模型研究加料单元数目对结果精度的影响。图 7(a)是裂尖采用一层加料单元的有限元模型,加料单元与常规单元通过过渡单元连接,图 7(b)是裂尖采用两层奇异单元的有限元模型。表 2 给出相应的计算结果。从表中可以

看出,增加裂纹尖端的加料单元数可以提高裂尖张开位移和能量释放率结果精度,而且对后者精度提高作用更明显一些。

6 结 论

(1)本文将加料有限元法应用于三维线粘弹性裂纹问题,该方法不仅可以获得裂尖附近位移,而且可以利用附加自由度获得裂纹能量释放率,从而延续了该方法一贯的优势,通过典型裂纹结构算例分析,验证了理论推导的正确性和方法的有效性;

(2)当加料单元边长达到裂纹长度的 $1/50 \sim 1/20$ 后,加料有限元法计算得到的结果精度明显高于常规有限元法,而且本文方法对网格不敏感;

(3)高斯积分点数、加料单元数对附加自由度项精度影响十分明显,在使用加料单元时,需要采用 $8 \times 8 \times 8$ 的高斯积分点数以及两层加料单元以保证能量释放率结果精度。

参考文献 (References)

[1] 蒙上阳. 基于粘弹性有限元方法的固体火箭发动机结构完整性分析[D]. 长沙:国防科学技术大学,2005.
MENG Shangyang. The structure integrity analysis of solid rocket motor based on viscoelastic finite element method [D]. Changsha: National University of Defense Technology, 2005. (in Chinese)

[2] Benzley S E. Representation of singularities with isoperimetric finite elements [J]. International Journal for Numerical Methods in Engineering, 1974, 8: 537 – 545.

[3] Stephane S P, Paul F J, Sherrill B B. A finite element analysis of the singular stress fields in anisotropic materials loaded in antiplane shear [J]. International Journal for Numerical Methods in Engineering, 1995, 38: 81 – 97.

[4] Stephane S P, Sherrill B B. Finite element evaluation of free – edge singular stress fields in anisotropic materials [J]. International Journal for Numerical Methods in Engineering, 1995, 38: 2225 – 2239.

[5] Stephane S P, Paul F J, Sherrill B B. Finite element analysis of singular stress fields in anisotropic materials loaded inplane [J]. International Journal of Solids Structures, 1995, 32: 571 – 591.

[6] Stephane S P, Sherrill B B. A finite element approach to three-dimensional singular stress states in anisotropic multi-material wedges and junctions [J]. International Journal of Solids and Structures, 1996, 33: 33 – 47.

[7] Stephane S P, Sherrill B B. Enriched finite elements for regions with multiple, interacting singular fields [J]. American Institute of Aeronautics and Astronautics Journal (AIAA Journal), 1996, 34(9): 1927 – 1933.

[8] Stephane S P, Sherrill B B. Enrichment of finite elements with numerical solutions for singular stress fields [J]. International Journal for Numerical Methods in Engineering, 1997, 40: 2693 – 2713.

[9] Bayram Y B, Nied H F. Enriched finite element-penalty function method for modeling interface cracks with contact [J]. Engineering Fracture Mechanics, 2000, 65: 541 – 557.

[10] Ayhan A O, Nied H F. Stress intensity factors for three-dimensional surface cracks using enriched finite elements[J]. International Journal for Numerical Methods in Engineering, 2002, 54: 899 – 921.

[11] Ayhan A O. Mixed mode stress intensity factors for deflected and inclined corner cracks in finite-thickness plates [J]. International Journal for Fatigue, 2007,29:305 – 317.

[12] Ayhan A O. Stress intensity factors for three-dimensional cracks in functionally graded materials using enriched finite elements [J]. International Journal of Solids and Structures, 2007,44: 8579 – 8599.

[13] Ayhan A O. Three-dimensional mixed-mode stress intensity factors for cracks in functionally graded materials using enriched finite elements [J]. International Journal of Solids Structures, 2009, 46: 796 – 810.

[14] Zhang C Y. Viscoelastic fracture mechanics [M]. Beijing: Science Press, 2006.

[15] 王自强, 陈少华. 高等断裂力学 [M]. 北京: 科学出版社, 2008.
WANG Ziqiang, CHENG Shaohua. Advanced fracture mechanics [M]. Beijing: Science Press, 2008. (in Chinese)

非定常 Euler 方程数值计算中高精度格式应用*

赵云飞¹, 刘 绪¹, 涂国华², 刘 伟¹, 毛枚良²

(1. 国防科技大学 航天与材料工程学院, 湖南 长沙 410073;

2. 空气动力学国家重点实验室, 四川 绵阳 621000)

摘 要: 采用五阶精度加权紧致非线性格式(WCNS)和非定常“双时间步”方法求解非定常 Euler 方程, 模拟 NACA0012 翼型强迫俯仰振动流场, 研究了高精度格式应用到非定常计算时“双时间步”方法物理时间步长、子迭代收敛判据、子迭代步数以及物理时间导数离散方法对计算精度和计算效率的影响。

关键词: 高精度格式; 非定常流动; WCNS 格式; Euler 方程; 数值模拟

中图分类号: V211.3 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0012-05

High-accuracy numerical method applied to
calculate unsteady Euler equations

ZHAO Yunfei¹, LIU Xu¹, TU Guohua², LIU Wei¹, MAO Meiliang²

(1. College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China;

2. State Key Laboratory of Aerodynamics, Mianyang 621000, China)

Abstract: By applying 5-order accuracy WCNS (Weighted Compact Nonlinear Schemes) to simulate forced-oscillation of NACA0012 aerofoil, some techniques of high-accuracy method in unsteady simulation were studied, such as inner iteration capability, inner iteration parameters of “dual time stepping” method, and discretization accuracy of time derivatives of Euler equations. The results show that, these coefficients have a remarkable effect on computation accuracy and CPU time when WCNS scheme is applied to unsteady flow simulation.

Key words: high-order accurate scheme; unsteady flows; WCNS schemes; Euler equations; numerical simulation

高精度、高分辨率格式概念自 20 世纪 80 年代由 Harten 提出以来一直被人们密切关注,其在理论和应用上的诸多难点也给研究人员带来了巨大挑战,时至今日,高精度数值格式研究仍然是 CFD 的前沿内容。高精度格式不仅在表面热流、摩阻、流动分离等需要正确模拟真实物理粘性的问题上具有广泛应用,而且在湍流直接数值模拟、大涡模拟、计算声学等对时间和空间具有多尺度特征的流动数值模拟上也具有迫切的应用背景。航空航天研究的需求不断推动着高精度格式快速发展,高精度格式在计算流体力学工作中发挥的作用正逐步增加。

目前,高精度格式应用于定常流动计算已较多,但针对非定常流动的研究尚不多见,国内外针对这方面的研究文献还不充分。实际上,非定常流动十分复杂,具有多频谱特征,常常伴随有多尺度的漩涡结构,对于非定常流场的这些精细结构,采用高精度、高分辨率的格式模拟有助于揭示复

杂流场的演化过程。在非定常计算中,Jameson^[1-2]发展的“双时间步”(Dual-time-step)方法在数值求解非定常 Euler/N-S 方程中得到了十分广泛的应用。“双时间步”方法涉及到子迭代收敛性、子迭代参数取值(物理时间步长、收敛判据、子迭代限制步数)等实际问题^[5],这对于高精度格式的应用来说十分重要。目前,高精度格式在这方面的计算经验还较少。另外,时间精度也是非定常计算的重要问题之一,在空间项采用高精度格式离散时,时间导数也要保证较高的离散精度。然而目前的隐式“双时间步”方法通常对时间导数只采用二阶格式离散,三阶或更高阶的离散方法研究还较为缺乏,因此这也是高精度非定常计算值得关注的问题。

针对上述问题,本文采用空间五阶精度的加权紧致非线性格式 WCNS^[3,6-8] 对非定常计算的这些关键基础问题进行了研究,包括非定常“双时间步”方法子迭代收敛性、子迭代参数选取、不

* 收稿日期:2011-06-24

基金项目:国家部委资助项目;国家自然科学基金资助项目(11172325,90716015)

作者简介:赵云飞(1988—),男,内蒙赤峰人,博士研究生,E-mail:zhaoyf08@126.com;

刘伟(通信作者),男,教授,博士,硕士生导师,E-mail:fishfather6525@sina.com

同时间离散方法对非定常计算的影响等。

1 控制方程

贴体坐标系下守恒形式的无量纲 Euler 方程为

$$\frac{\partial \bar{Q}}{\partial t} + \frac{\partial \bar{E}}{\partial \xi} + \frac{\partial \bar{F}}{\partial \eta} + \frac{\partial \bar{G}}{\partial \zeta} = 0 \quad (1)$$

式中 \bar{Q} 为守恒变量, \bar{E} 、 \bar{F} 、 \bar{G} 为无粘通量。

2 数值方法

2.1 空间离散方法

Euler 方程对流项采用五阶精度原始变量型 WCNS 格式离散,以 ξ 方向为例(网格间距为 h):

$$\begin{aligned} \bar{E}'_j &= \frac{75}{64h} (\bar{E}_{j+\frac{1}{2}} - \bar{E}_{j-\frac{1}{2}}) - \\ &\frac{25}{384h} (\bar{E}_{j+\frac{3}{2}} - \bar{E}_{j-\frac{3}{2}}) + \frac{3}{640h} (\bar{E}_{j+\frac{5}{2}} - \bar{E}_{j-\frac{5}{2}}) \end{aligned} \quad (2)$$

设 U 为原始变量,那么式中的 $\bar{E}_{j\pm\frac{1}{2}} = E(\tilde{U}_{j\pm\frac{1}{2}})$ 代表单元边界的数值通量,本文采用通量分裂方法计算:

$$\bar{E}_{j\pm\frac{1}{2}} = E^-(\tilde{U}_{Rj\pm\frac{1}{2}}) + E^+(\tilde{U}_{Lj\pm\frac{1}{2}}) \quad (3)$$

正负通量 E^\pm 由 Steger-Warming 分裂方法得到,其中 $\tilde{U}_{Rj\pm\frac{1}{2}}$ 和 $\tilde{U}_{Lj\pm\frac{1}{2}}$ 为单元边界上的原始变量,通过高阶非线性插值求得。WCNS 边界格式及相关方法具体参见文献[10]。

2.2 时间离散方法

本文采用 LU-SGS + “双时间步”(Dual-time-step)方法离散时间导数项:

$$L(\Delta Q^{p+1}) = RHS^p \quad (4)$$

$$RHS^p = - \left[J^{-1} \frac{\partial(Q^p)}{\partial t} + R(Q^p) \right]$$

式(4)中,上角标 p 表示子迭代步数。左端项表示对虚拟时间导数的隐式 LU-SGS 方法处理^[9],右端项 RHS^p 包含非定常方程时间导数项 $\partial(Q^p)/\partial t$ 和空间导数项 $R(Q^p)$,空间导数项采用 WCNS 格式离散。本文分别用时间一阶精度、二阶精度和三阶精度的后差格式对 $\partial(Q^p)/\partial t$ 离散:

时间导数一阶离散:

$$RHS^p = - \left[J^{-1} \frac{Q^p - Q^n}{\Delta t} + R(Q^p) \right] \quad (5a)$$

时间导数二阶离散:

$$RHS^p = - \left[J^{-1} \frac{3Q^p - 4Q^n + Q^{n-1}}{2\Delta t} + R(Q^p) \right] \quad (5b)$$

时间导数三阶离散:

$$RHS^p = - \left[J^{-1} \frac{11Q^p - 18Q^n + 9Q^{n-1} - 2Q^{n-2}}{6\Delta t} + R(Q^p) \right] \quad (5c)$$

3 计算结果及分析

3.1 计算模型

NACA0012 翼型强迫俯仰振动定义为攻角随时间正弦变化:

$$\alpha = \alpha_0 + \alpha_m \sin(kt) \quad (6)$$

式中 α 为瞬时攻角, α_0 为起始攻角, α_m 为振幅, t 是无量纲物理时间 $t = (\tilde{t} \cdot \tilde{V}_\infty) / \tilde{L}_{ref}$, \tilde{L}_{ref} 和 \tilde{V}_∞ 分别是控制方程无量纲化的特征长度和特征速度, k 是减缩频率。本文所有计算均采用此强迫振动模型,基本计算条件包括: $Ma = 0.755$, $\alpha_0 = 0.016^\circ$, $T_\infty = 300K$, 振幅 $\alpha_m = 2.51^\circ$, 减缩频率 $k = 0.1628$ 。

3.2 “双时间步”方法对高精度非定常计算的若干影响因素研究

本节考查高精度格式应用到非定常计算时,“双时间步”方法物理时间步长、子迭代收敛判据以及最大子迭代步数对非定常计算的影响,采用第2节给出的数值方法求解非定常 Euler 方程数值模拟 NACA0012 翼型强迫俯仰振动,时间导数采用(5b)式的二阶格式离散。

3.2.1 物理时间步长的影响

采用空间五阶精度 WCNS 格式在 6 个不同物理时间步长 dt 下(0.01、0.02、0.05、0.1、0.2、0.5)数值模拟 NACA0012 翼型强迫俯仰振荡,子迭代收敛判据统一设定为 $Tol = 0.01$ (子迭代收敛判据定义见文献[10]),子迭代均达到收敛判据后跳出。图1是不同时间步长得到的升力迟滞曲线对比,由于非定常迟滞作用,曲线表现出迟滞环形态。时间步长小于 0.1 的 4 条曲线基本是重合的,并与参考文献计算结果基本吻合^[4]。当时时间步长较大取 0.2 时曲线有了少许偏离,而到 0.5 时曲线偏离程度已经很大,在最大攻角处曲线有明显变形,这说明时间步长过大后计算精度得不到满足。从曲线在最大攻角处的放大图可以看出,随着物理时间步长 dt 的减小,曲线呈现单向收敛的趋势。对于本算例, $dt = 0.1$ 是一个恰当的选择,在保证计算精度的同时,时间步长相对较大,从而减少一个振动周期的计算步数,提高计算效率。

表1列出了第3个振动周期结束时不同时间步长总的计算 CPU 时间对比(以 $dt = 0.1$ 的 CPU

用时为单位折算),能够看出时间步长增大(减小)的确缩短(增加)了计算时间,但这种缩短(增加)并不是线性的,原因是时间步长增大(减小)在减少(增加)计算步数的同时却增加(减少)了子迭代收敛步数。

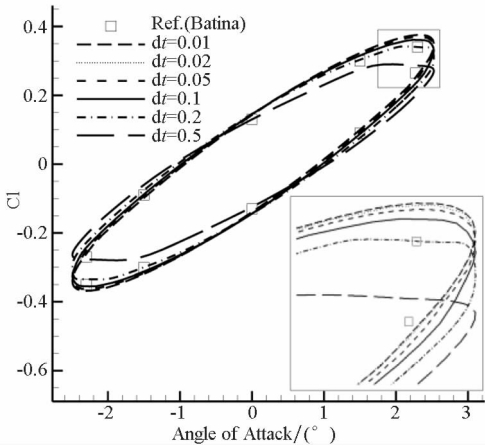


图 1 物理时间步长对计算精度的影响

Fig. 1 Influence of the physical time-step

表 1 不同物理时间步长的计算 CPU 时间对比

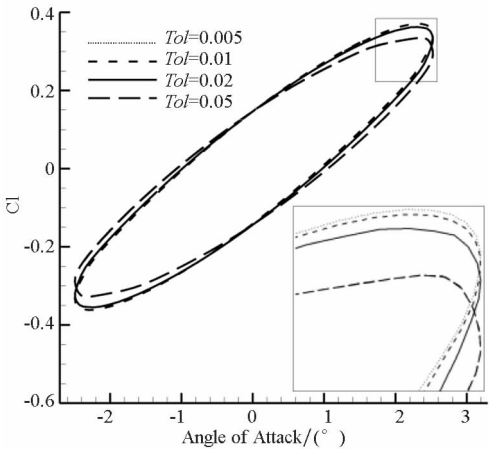
Tab. 1 Comparison of CPU time with different physical time-step

物理时间步长 dt	0.01	0.02	0.05	0.1	0.2	0.5
CPU 时间	3.54	2.6	1.52	1	0.63	0.31

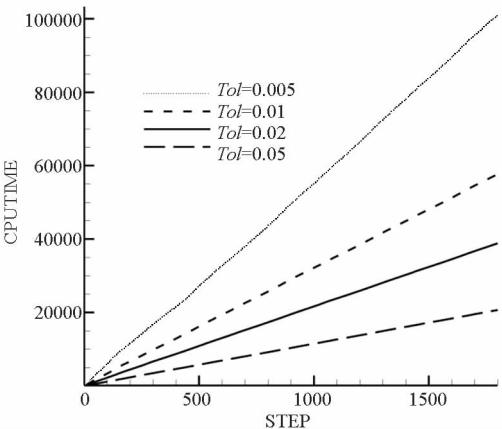
3.2.2 子迭代收敛判据的影响

图 2 是收敛判据 Tol 分别取 0.005、0.01、0.02、0.05 的计算结果,计算物理时间步长统一取 0.05,子迭代不给最大步数限制,使之达到收敛判据后才跳出。由图 2(a)可见,收敛判据取 0.01 和 0.005 的迟滞环几乎是重合的,也就是说收敛判据取 0.01 可以满足精度要求,再往小取就没什么意义了。相反地,收敛判据较大时,取 0.02 的曲线较 0.01 稍微有了些偏离,但还可以接受,而到了 0.05 时,迟滞曲线在最大攻角处有了明显变形,不能满足计算精度要求。

图 2(b)是不同收敛判据下的计算 CPU 时间比较,横轴是计算物理迭代步数,纵轴是计算 CPU 时间。可以看出,随着收敛判据的减小,计算 CPU 时间在不断增加,到计算结束 (step = 1800) 时,收敛判据 0.005 用了 100000 个 CPU 时间,收敛判据 0.01 只用了不到 60000 的 CPU 时间,几乎少了一半,而两者之间的计算精度却差异不大。从以上的分析中可以看到,采用高精度格式进行非定常计算时,子迭代收敛判据对计算精度和效率有重要影响。



(a) 对迟滞环计算精度的影响



(b) 对计算 CPU 时间的影响

图 2 收敛判据的影响

Fig. 2 Influence of inner iteration tolerance (Tol)

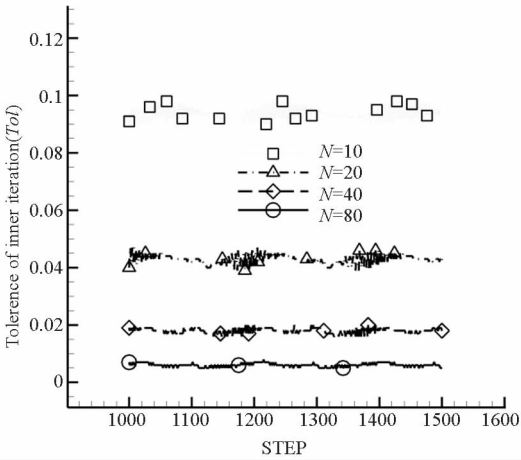
3.2.3 最大子迭代步数的影响

设定最大子迭代步数可以保证子迭代在无法达到收敛判据时同样能够跳出子迭代,以免计算陷入死循环,这是目前非定常计算常采用的方法。本节在最大子迭代步数分别取 10、20、40、80 的情况下采用 WCNS 格式模拟 NACA0012 翼型强迫俯仰振荡,时间步长统一取 0.1,并保证子迭代都达到最大子迭代步数才跳出,这样能够单方面研究子迭代步数对计算结果的影响。

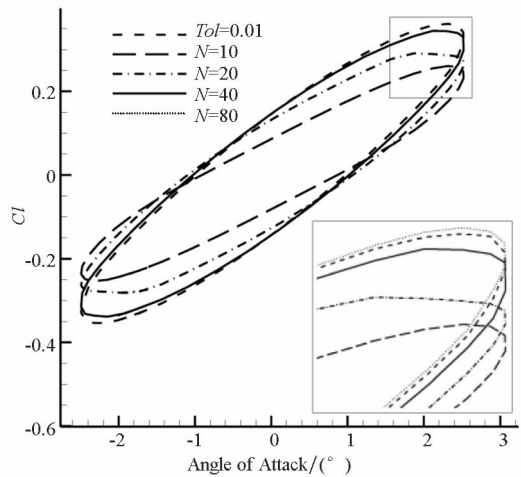
图 3(a)是 4 种不同子迭代步数的收敛水平比较。子迭代步数限制在 10 步时,对应子迭代残差收敛标准在 0.1 附近,随着子迭代步数增多,子迭代残差也收敛得越来越小,子迭代步数为 40 时,大致对应 0.02 的残差标准,子迭代步数限制在 80 步时,残差已收敛到 0.01 左右。

图 3(b)是不同子迭代步数的升力迟滞环的比较。图中 $Tol = 0.01$ 与 $N = 80$ 的曲线几乎重合,因为它们的残差收敛水平一样都是 0.01 左右(参考图 3(a))。随着子迭代步数的减少,残差收敛水平越来越差, $N = 20$ 和 $N = 10$ 的曲线已偏

离太大,不能满足计算精度。上述研究说明,在对实际问题计算时,子迭代步数限制标准对计算结果的影响是一个很重要的问题。



(a) 不同子迭代步数对应的收敛水平



(b) 对迟滞环计算精度的影响

图3 最大子迭代步数的影响

Fig. 3 Influence of the times of inner iteration

3.3 不同时间离散方法对高精度非定常计算的影响

空间项采用高精度 WCNS 格式离散时,时间导数的离散同样要保证较高的精度。对时间导数项分别用式(5a)、(5b)、(5c)给出的时间一阶、二阶、三阶后差格式离散,数值模拟 NACA0012 翼型强迫俯仰振荡,考查时间离散方法对计算结果的影响。

图4是在不同物理时间步长 dt 和收敛标准 Tol 下的计算结果。从图4(a)的升力迟滞环来看,控制方程时间导数项离散方法不同对计算结果的影响并不明显,三种离散方法迟滞环几乎重合。然而从俯仰力矩迟滞环来看,如图4(b)和4(c),三阶离散方法的计算曲线随 dt 增大出现不同程度的非物理振荡, $dt=0.1$ 振荡较弱, $dt=0.2$

时振荡最为明显。如果保持图4(c)的其他计算条件不变,只把子迭代收敛标准提高到0.005,如图4(d),振荡现象则削弱很多。可见缩小物理时间步长或提高子迭代收敛标准能抑制时间高阶离散所引起的振荡问题(但会增加计算时间)。

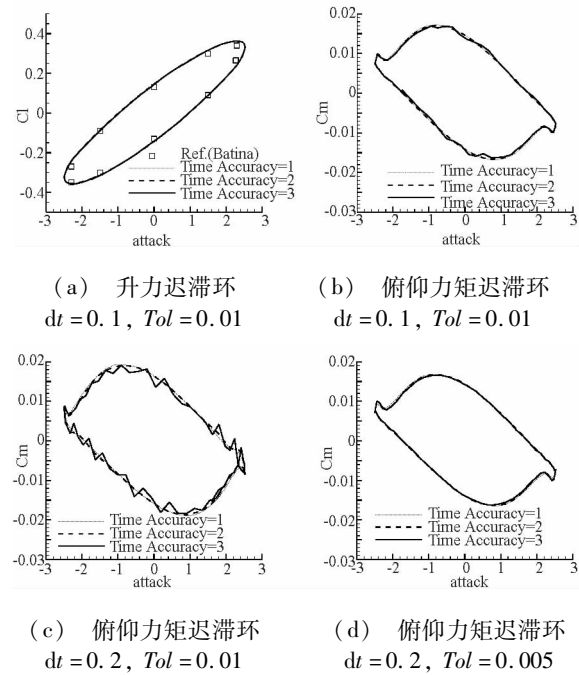


图4 时间离散方法不同对计算结果的影响

Fig. 4 Influence of time discretization accuracy on numerical results

另外,比较3种离散方法的计算CPU时间,由表2可见,一阶时间离散方法用时稍长一些,二、三阶时间离散方法用时较短。通过分析,其原因是在相同的收敛判据下($Tol=0.01$),一阶离散方法计算子迭代步数较多,而二、三阶离散子迭代步数较少,这从图5可以得到验证(横轴为物理推进步数,纵轴为子迭代步数)。

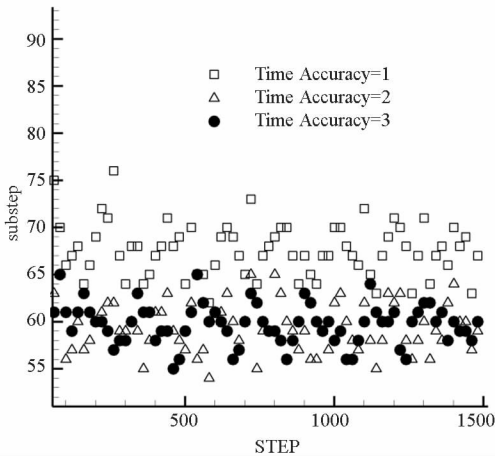


图5 时间离散方法不同时子迭代收敛步数的比较

Fig. 5 Times of inner iteration for three time discretization methods ($dt=0.1$, $Tol=0.01$)

总体来看,在本文空间采用五阶精度 WCNS 格式离散条件下,时间导数项采用二阶格式离散在计算精度、稳定性以及计算效率上都较为理想。

表 2 不同时间离散方法的计算 CPU 时间对比

Tab. 2 Comparison of CPU time with different time
discretion accuracy($dt=0.1$, $Tol=0.01$)

时间离散方法	一阶离散	二阶离散	三阶离散
CPU 时间	1.11	1.0	1.03

4 结 论

本文基于空间五阶精度格式 WCNS,研究了高精度格式应用到非定常计算时,“双时间步”方法子迭代参数选取、子迭代收敛性和控制方程时间导数不同离散精度等问题。对高精度计算方法来说,物理时间步长过大会降低非定常计算精度,取得过分小会使计算周期延长,从而增加计算时间;时间步长增大同样会降低子迭代收敛能力。子迭代要达到一定的收敛水平才能满足计算精度的需要,超过这个限度后计算精度的提升不大,反而会使计算时间增加。基于空间五阶精度 WCNS 格式的时间离散方法研究表明:一阶离散方法计算收敛能力较差,计算效率稍低;三阶离散方法出现非物理振荡(提高子迭代收敛水平能够削弱这种振荡);而时间二阶离散方法从精度、效率等方面总体计算表现较好。

参考文献 (References)

[1] Jameson A. Time dependent calculations using multigrid with applications to unsteady flows past airfoils and wings [R].

AIAA Paper,1991;1596.

[2] Hsu J M, Jameson A. An implicit-explicit hybrid scheme for calculating complex unsteady flows [R]. AIAA Paper, 2002;0714.

[3] Deng X G, Mao M L. Weighted compact high-order nonlinear schemes for the Euler equations [R]. AIAA Paper, 1997;1941.

[4] Batina J T. Unsteady Euler airfoil solutions using unstructured dynamic meshes[R]. AIAA Paper,1989;0115.

[5] 赵海洋,刘伟,任兵. 双时间隐式方法求解非定常绕流的相关问题[J]. 国防科技大学学报, 2005, 27(3): 25 - 29. ZHAO Haiyang, LIU Wei, REN Bing. The unsteady flow field calculation with a dual-time stepping method[J]. Journal of National University of Defense Technology, 2005, 27(3): 25 - 29. (in Chinese)

[6] Deng X G, Zhang H X. Developing high-order weighted compact nonlinear schemes [J]. Journal of Computational Physics, 2000, 165: 22 - 44.

[7] Deng X G, Mao M L, Tu G H, et al. Extending weighted compact nonlinear schemes to complex grids with characteristic based interface conditions[J]. AIAA Journal 2010, 48(12): 2840 - 2851.

[8] Deng X G, Mao M L, Tu G H, et al. Geometric conservation law and applications to high-order finite difference schemes with stationary grids[J]. Journal of Computational Physics, 2011, 230: 1100 - 1115.

[9] Yoon S, Jameson A. Lower-upper symmetric Gauss-Sediel method for the Euler and Navier-stokes equations[J]. AIAA Journal, 1988, 26(9): 1025 - 1026.

[10] 张毅锋. 高阶精度格式(WCNS)加速收敛和复杂流动数值模拟的应用研究[D]. 绵阳:中国空气动力研究与发展中心,2007. ZHANG Yifeng. Investigations of convergence acceleration and complex flow numerical simulation for high-order accurate scheme (WCNS) [D]. Mianyang: China Aerodynamics Research and Development Center, 2007. (in Chinese)

高超声速滑翔飞行器倾斜转弯分析及控制系统设计*

陈小庆,侯中喜,刘建霞
(国防科技大学 航天与材料工程学院,湖南 长沙 410073)

摘要: 倾斜转弯技术是高超声速滑翔飞行器控制的一个重要发展方向。针对高超声速滑翔飞行器倾斜转弯技术开展研究。以平衡滑翔弹道为参考弹道,分析了转弯半径、下降高度、倾侧角等参数之间的关系,提出在设计高超声速滑翔飞行器制导控制指令时,应综合考虑不同高度速度下的控制能力约束。根据奇异摄动理论将动力学系统的受控状态变量分为快变量和慢变量两部分,运用轨迹线性化方法设计了控制系统。仿真结果表明,设计的控制器具有良好的控制性能,但随着高度的增加,控制指令应结合实际控制能力,以完成对飞行器的姿态控制。

关键词: 高超声速滑翔飞行器; 倾斜转弯; 轨迹线性化; 飞行控制系统
中图分类号: V44 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0017-07

Analysis and controller design of bank-to-turn system for hypersonic gliding vehicle

Chen Xiaoqing, Hou Zhongxi, Liu Jianxia
(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Bank-to-turn (BTT) technology is an important development direction of control technology for hypersonic gliding vehicles. This research focuses on the control system of bank-to-turn for the hypersonic gliding vehicle. By taking the equilibrium gliding trajectory as the reference trajectory, it analyzed the relationship between turn radius, falling height, and bank angle. The analysis shows that when designing the guiding and control command, the flight conditions and the limitation of the control ability need to be considered. Then the dynamics were separated into fast and slow variables in terms of singular perturbation theory. Furthermore, the trajectory linearization control (TLC) method was applied to design the control system. Simulation results show that the controller is robust and the command should be adjusted with the flight conditions to control the vehicle efficiently.

Key words: hypersonic gliding vehicle; bank-to-turn; trajectory linearization control; flight control system

再入滑翔飞行器具有远程快速到达、高速精确投送、大范围区域覆盖、纵/横向机动性好等优点。与传统再入飞行器相比,再入滑翔飞行器充分利用其高升阻比的气动外形在临近空间内飞行,采用滑翔和跳跃的组合弹道,不但能够增加飞行时间,提高射程,而且还能有效地降低壁面最大热流密度。

控制系统是再入滑翔飞行器重要分系统之一,是飞行器安全飞行,完成既定任务的重要保证。当前针对再入飞行器的控制研究多以空天飞行器为对象,文献[1]以一概念性升力体高超声速飞行器为对象,对其再入段进行了研究,建立了数学模型,并研究了再入段的控制系统研究方法。文献[2]利用非线性动态逆技术设计了高超声速飞行器快慢回路的控制律,并针对飞行器在高超

声速下的倾斜转弯运动进行了仿真。文献[3]建立了空天飞行器再入飞行条件下的6自由度模型;文献[4]则研究了基于模糊理论的空天飞行器鲁棒自适应控制,文献[5]研究了基于Terminal滑模的空天飞行器自适应鲁棒控制,文献[6]则基于轨迹线性化方法研究了空天飞行器的控制方法。

1 数学模型

1.1 动力学模型

由于以姿态角作为姿态控制输出时高超声速飞行器再入模式的动力学方程呈现非最小相位行为,因此姿态动力学方程一般以攻角、倾斜角和侧滑角作为输出。数学模型的具体推导可参考文献[7-8],考虑到飞行器为无动力滑翔,其动力学方

* 收稿日期:2011-03-01
基金项目: 国家自然科学基金资助项目(90916016);湖南省优秀研究生创新基金资助项目
作者简介:陈小庆(1982—),男,江苏泰兴人,博士,E-mail: gkbatchelor@gmail.com;
侯中喜(通信作者),男,教授,博士,博士生导师,E-mail: cn_hzx@sina.com.cn

程如下:

$$\begin{cases}
 \dot{\alpha} = q - \tan\beta(p\cos\alpha + r\sin\alpha) + \frac{-L + mg\cos\gamma\cos\mu}{mV\cos\beta} \\
 - \frac{V\cos\gamma}{r\cos\beta}(\cos\mu + \sin\mu\cos\gamma\sin\chi\tan\varphi) \\
 \dot{\beta} = -r\cos\alpha + p\sin\alpha + \frac{Y\cos\beta + mg\sin\mu\cos\gamma}{mV} \\
 - \frac{V\cos\gamma}{r}(\sin\mu + \sin\chi\tan\varphi\cos\mu\cos\gamma) \\
 \dot{\mu} = \sec\beta(p\cos\alpha + r\sin\alpha) - \frac{g}{V}\tan\beta\cos\mu\cos\gamma \\
 + \frac{L(\tan\beta + \tan\gamma\sin\mu) + Y\tan\gamma\cos\beta\cos\mu}{mV} \\
 + \frac{V\cos\gamma}{r}[\tan\beta\cos\mu + \sin\chi\tan\varphi(\sin\gamma + \tan\beta\sin\mu\cos\gamma)] \\
 \dot{p} = I_{pq}^p pq + I_{qr}^p qr + g_n^p n + g_l^p l \\
 \dot{q} = I_{pp}^q p^2 + I_{rr}^q r^2 + I_{pr}^q pr + g_m^q m \\
 \dot{r} = I_{qr}^r qr + I_{pq}^r pq + g_l^r l + g_n^r n
 \end{cases} \quad (1)$$

其中:

$$\begin{aligned}
 I_{pq}^p &= \frac{I_{xz}(I_{xx} - I_{yy} + I_{zz})}{I_{xx}I_{zz} - I_{xz}^2}, & I_{pp}^q &= \frac{-I_{xz}}{I_{yy}}, & I_{rr}^q &= \frac{I_{xz}}{I_{yy}}, \\
 I_{qr}^p &= \frac{I_{zz}(I_{yy} - I_{zz}) - I_{xz}^2}{I_{xx}I_{zz} - I_{xz}^2}, & I_{pr}^q &= \frac{(I_{zz} - I_{xx})}{I_{yy}}, \\
 I_{pq}^r &= \frac{I_{xx}^2 - I_{xx}I_{yy} + I_{xz}^2}{I_{xx}I_{zz} - I_{xz}^2}, & I_{qr}^r &= \frac{-I_{xz}(I_{xx} - I_{yy} + I_{zz})}{I_{xx}I_{zz} - I_{xz}^2}, \\
 g_l^p &= \frac{I_{zz}}{I_{xx}I_{zz} - I_{xz}^2}, & g_n^p &= \frac{I_{xz}}{I_{xx}I_{zz} - I_{xz}^2}, & g_m^q &= \frac{1}{I_{yy}}, \\
 g_l^r &= \frac{I_{xz}}{I_{xx}I_{zz} - I_{xz}^2}, & g_n^r &= \frac{I_{xx}}{I_{xx}I_{zz} - I_{xz}^2}.
 \end{aligned}$$

α, β, μ 分别为攻角、侧滑角和倾侧角, p, q, r 分别为飞行器滚转、俯仰和偏转角速度; $I_{xx}, I_{yy}, I_{zz}, I_{xz}$ 为飞行器转动惯量分量; l, m, n 为飞行器所受的三轴力矩。 L, Y 分别为飞行器所受到的升力和侧向力, m 为飞行器质量, V 为速度。

1.2 气动参数模型

本文中气动参数采用乘波构型气动力数据拟合得到, 设计的气动控制面和 HTV-1 类似, 共有 6 个控制面, 如图 1 所示。本文所研究的滑翔飞行器基于乘波构型设计, 气动参数为乘波构型气动参数。气动力和气动力矩计算表达式为

$$\begin{aligned}
 L &= \bar{q}SC_D, D = \bar{q}SC_L, Y = \bar{q}SC_Y, \\
 l &= \bar{q}bSC_l, n = \bar{q}bSC_n + x_{gp}Y, \\
 m &= \bar{q}cSC_m - x_{gp}(D\sin\alpha + L\cos\alpha)
 \end{aligned}$$

其中:

$$\begin{aligned}
 C_L &= C_{L0} + C_{L,\alpha}\alpha \\
 C_D &= C_{D0} + C_{D,\alpha}\alpha + C_{D,\alpha^2}\alpha^2 \\
 C_Y &= C_{Y,\beta}\beta
 \end{aligned} \quad (2)$$

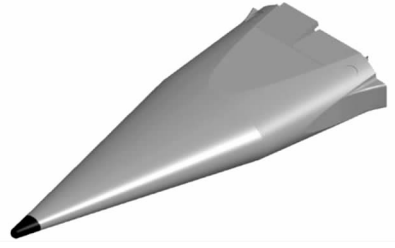


图 1 HTV-1 气动布局

Fig. 1 Aerodynamic design of HTV-1

x_{gp} 为质心到参考力矩中心(压心)的距离, 质心在参考力矩中心后为正。控制面的参数通过估算方法得到。假设飞行器共有 N 个气动控制面, 则

$$\begin{aligned}
 C_l &= C_{l,\beta}\beta + \sum_{i=1}^N C_{l,\delta_i}\delta_i + C_{l,p}\frac{pb}{2V} + C_{l,r}\frac{rb}{2V} \\
 C_m &= C_{m,\alpha} + \sum_{i=1}^N C_{m,\delta_i}\delta_i + C_{m,q}\frac{qc}{2V} \\
 C_n &= C_{n,\beta}\beta + \sum_{i=1}^N C_{n,\delta_i}\delta_i + C_{n,p}\frac{pb}{2V} + C_{n,r}\frac{rb}{2V}
 \end{aligned} \quad (3)$$

δ_i 为各控制面的舵偏角。由于在再入初级阶段, 大气稀薄, 气动舵面不能提供足够的力来控制飞行器的姿态, 除了气动控制面外, 再入滑翔飞行器还应包含喷流反作用控制系统(RCS)。随着飞行高度的降低, 动压逐渐增大, 逐渐过渡到采用气动控制面进行姿态控制^[3, 9], 本文主要研究大气层内的姿态控制, 不考虑 RCS 的建模及控制。

2 倾斜转弯分析

高超声速滑翔飞行器按照设计要求, 需要具备一定的侧向机动能力。采用倾斜转弯(BTT: Bank-To-Turn)技术进行机动较侧滑转弯具有较大的优势, 因此, 本文针对倾斜转弯相关参数进行了分析。

文献[10]通过研究最优滑翔轨迹指出: 飞行器在滑翔过程中, “除了要求对升力控制量进行初段和末段的调整, 在滑翔飞行的主要阶段, 接近于最大升阻比的滑翔是最优的滑翔”; 即在初始阶段需通过控制飞行器的姿态使其满足平衡滑翔条件, 从而获得较优的性能。这里以平衡滑翔弹道为参考弹道进行分析。

平衡滑翔条件下速度和密度(高度)的关系可由(4)式描述^[11]:

$$\rho = \frac{2m(g r_0 - V^2)}{S \cdot C_L \cdot r_0 \cdot V^2} \quad (4)$$

其中 S 为参考面积, g 为重力加速度, r_0 为地心距。倾侧角为 μ 时, 飞行器受力如图 2 所示。

假设飞行器倾斜转弯过程受力处于平衡

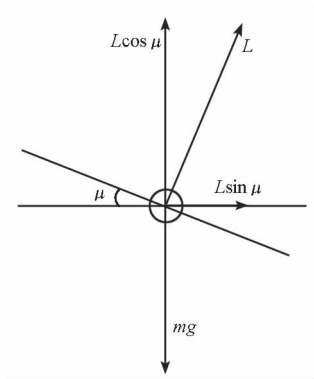


图 2 倾斜转弯受力分析
Fig. 2 Force analysis of bank-to-turn

状态:

$$\begin{aligned} L \cos \mu &= mg \\ L \sin \mu &= m \frac{V^2}{R} \end{aligned} \quad (5)$$

R 表示飞行器倾斜转弯时的转弯半径,令无量纲速度 $u = \frac{V^2}{gr_0}$,此时飞行器所处的高度对应的密度为

$$\rho' = \frac{2m(1-u)}{S \cdot C_L \cdot r_0 \cdot u \cdot \cos \mu} \quad (6)$$

倾侧角度为 μ 时,为保持竖直方向力平衡,飞行器下降高度对应的密度变化量

$$\Delta \rho = \frac{2m(1-u)}{S \cdot C_L \cdot r_0 \cdot u} \left(\frac{1}{\cos \mu} - 1 \right) \quad (7)$$

则在指数密度模型假设下,飞行器下降高度为

$$\Delta h = -\frac{1}{B} \ln(\cos \mu) \quad (8)$$

其中 B 为大气模型参数,上式表明,飞行器平衡滑翔时,倾斜转弯导致飞行器下降高度与飞行器气动参数、飞行速度等无关,只和倾侧角相关,关系如图 3 所示。

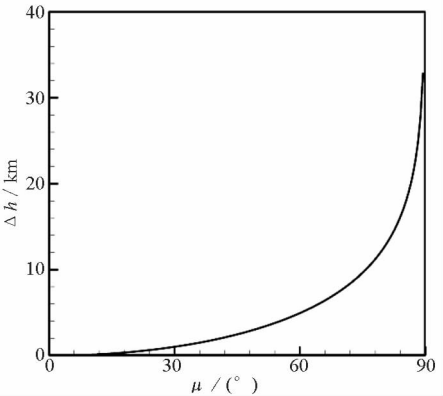


图 3 下降高度随倾侧角变化关系
Fig. 3 Decent height vs. bank angle

由(5) 式可知,转弯半径和倾侧角之间关系为

$$R = \frac{V^2}{g \tan \mu} = \frac{ur_0}{\tan \mu} \quad (9)$$

结合(6) 式可得转弯半径

$$R = \frac{2m \cdot r_0}{2m \cdot \tan \mu + \rho \cdot S \cdot C_L \cdot r_0 \cdot \sin \mu} \quad (10)$$

即转弯半径由来流状态、飞行器相关参数以及倾侧角共同决定。转弯半径随着升力系数的增加而降低。固定飞行器质量和升力系数的情况下,转弯半径和飞行高度之间的关系如图 4 所示。可以看出,转弯半径随着高度的降低而减小,相同高度时,转弯半径随着倾侧角的增大而减小。

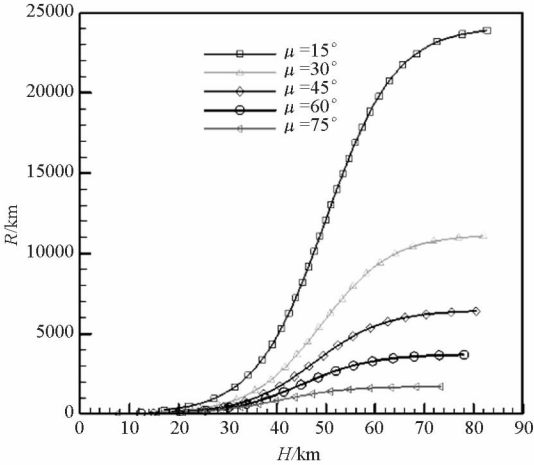


图 4 转弯半径随高度变化曲线
Fig. 4 Turn radius vs. height

只考虑滚转通道,假设高超声速条件下通过图 1 中的两侧副翼差动获得滚转力矩,则由动量矩方程:

$$I \frac{d\omega}{dt} = M \quad (11)$$

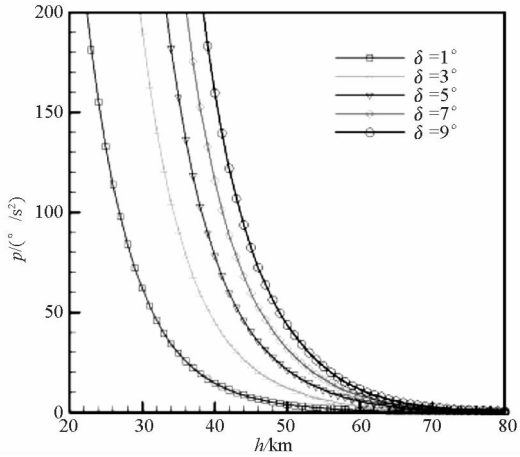
$$\frac{d\omega_x}{dt} = \frac{M}{I_{xx}} = \frac{\Delta p \cdot S \cdot L}{I_{xx}} \quad (12)$$

滚转角加速度主要由控制面上下表面的压强差 Δp 决定,而这取决于来流状态参数,图 5 为不同条件(相同马赫数,不同差动角;相同差动角,不同马赫数)下滚转角加速度随高度的变化曲线。

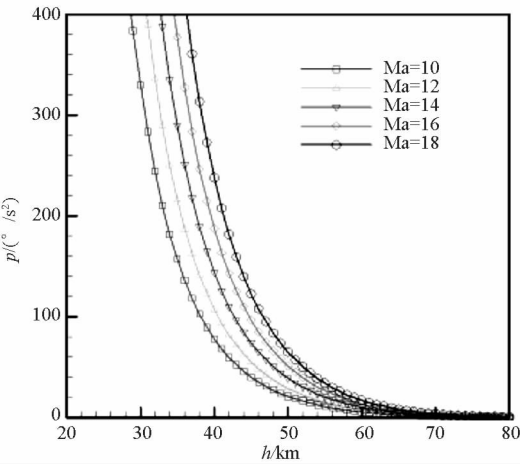
相同马赫数下,滚转角加速度随高度的增加而下降很快,舵面差动角度为 9° 时,在 40km 高度,滚转角加速度达到 $160^\circ/\text{s}^2$,而在 80km 高度,其滚转角加速度仅为 $1^\circ/\text{s}^2$ 。不同马赫数条件下,滚转角加速度随高度的变化呈现同样的趋势。故针对倾侧角的控制能力随着高度的增加而快速下降,在设计制导控制指令时,应综合考虑不同高度速度下的滚转角加速度约束。

3 控制系统设计

基于动力学方程式(1) 设计飞行控制系统。



不同舵面差动角 ($Ma = 10$)



不同马赫数 ($\delta = 5^\circ$)

图 5 滚转角加速度随高度变化曲线

Fig. 5 Roll angular acceleration vs. height

令姿态角 $\Omega = (\alpha, \beta, \mu)^T$, 姿态角速度 $\omega = (p, q, r)^T$, δ 为控制变量, 控制方程可改写成如下形式

$$\dot{\Omega} = f_s + g_{s1}\omega + g_{s2}\delta \quad (13)$$

$$\dot{\omega} = f_f + g_f M_C \quad (14)$$

M_C 为控制力矩, 其表达式为

$$M_C = g_{f,\delta}\delta \quad (15)$$

$g_{f,\delta}$ 为由控制面参数决定的矩阵。

3.1 快慢回路假设

高超声速滑翔飞行器的操纵面偏转所产生的操纵力比操纵力矩的作用小得多, 即舵面偏角 δ 的改变对姿态角加速度 $[\dot{p}, \dot{q}, \dot{r}]^T$ 有显著的影响, 而对 $[\dot{\alpha}, \dot{\beta}, \dot{\mu}]^T$ 的影响要小的多, 但 $[p, q, r]^T$ 的变化对 $[\dot{\alpha}, \dot{\beta}, \dot{\mu}]^T$ 却有很大影响, 因此称 $\omega = [p, q, r]^T$ 为系统快状态, 称 $\Omega = [\alpha, \beta, \mu]^T$ 为慢状态。基于这一事实, 可以利用奇异摄动理论, 为高超声速滑翔飞行器设计时标分离的快慢回路控制

器^[12-13]。忽略式(13)中的 g_{s2} , 则有

$$\dot{\Omega} = f_s + g_{s1}\omega \quad (16)$$

慢回路的设计目标是根据期望的制导指令 Ω_c 产生慢回路控制输入 ω_c , 并作为快回路的期望值, 用于设计期望的控制力矩 M_C 。为了保证设计的有效性, 必须要求快回路带宽足够大, 一般为慢回路的 3 ~ 5 倍。通过仿真可以检验, 忽略 g_{s2} 对受控状态的影响是较小的。高超声速滑翔飞行器快慢回路飞行控制系统结构如图 6 所示。

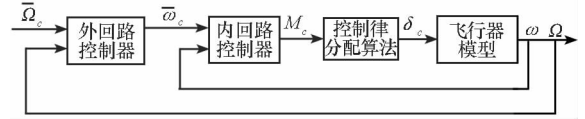


图 6 控制器结构

Fig. 6 Controller structure

3.2 轨迹线性化

轨迹线性化控制^[14] (Trajectory Linearization Control, TLC) 是 20 世纪 90 年代发展起来的一种非线性控制方法, 具有很好的非线性跟踪和解耦控制能力。其设计思想是将跟踪问题变成一个沿着标称轨迹的非线性调节问题, 将渐进的跟踪控制利用一个 2 自由度的控制器来实现: (1) 一个动态的被控对象输入输出的逆映射, 用以根据期望的系统输出来产生标称的控制输入, 由于理想条件下的精确逆系统难以获得, 所以采用伪逆系统, 这是一个开环控制器; (2) 一个跟踪误差稳定调节器, 用以解决伪逆、初始条件和系统不确定等因素的影响, 并使系统具有一定的响应特性, 这是一个闭环控制器, 具体控制结构如图 7 所示。

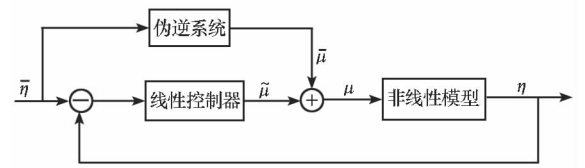


图 7 TLC 控制器

Fig. 7 TLC controller

3.3 标称指令计算

根据 TLC 方法的设计思想和步骤, 首先计算飞行器快慢回路的标称指令。由于慢回路标称指令为飞行器的制导指令, 即 $\bar{\Omega} = \Omega_c$, 因此不难求得标称的姿态角速率 $\bar{\omega}$ 为

$$\bar{\omega} = \bar{g}_{s1}^{-1}(\dot{\bar{\Omega}} - \bar{f}_s) \quad (17)$$

对于 $\bar{\omega}$ 来说, 它又成为快回路的标称指令, 同理求得标称的快回路控制力矩如下

$$\bar{M}_C = \bar{g}_f^{-1}(\dot{\bar{\omega}} - \bar{f}_f) \quad (18)$$

为了保证系统的因果性, $\dot{\bar{\Omega}}, \dot{\bar{\omega}}$ 由 $\bar{\Omega}, \bar{\omega}$ 经过如下伪微分器求得

$$G_{i,diff} = \frac{\omega_{i,diff}s}{s + \omega_{i,diff}}, i = 1, 2 \quad (19)$$

其中 $\omega_{i,diff}$ 为低通滤波器的带宽, 选取时应保证给定的标称指令可以通过, 同时又可以抑制高频噪音^[15]。

3.4 控制器设计

TLC 方法通过定义增广变量设计出具有比例积分形式的线性时变调节器来提高闭环系统的性能和鲁棒性。针对快慢回路有如下增广的系统状态向量定义:

$$\begin{aligned} x_1 &= \left[\int \alpha dt, \alpha, \int \beta dt, \beta, \int \mu dt, \mu \right]^T \\ \bar{x}_1 &= \left[\int \bar{\alpha} dt, \bar{\alpha}, \int \bar{\beta} dt, \bar{\beta}, \int \bar{\mu} dt, \bar{\mu} \right]^T \\ x_2 &= \left[\int p dt, p, \int q dt, q, \int r dt, r \right]^T \\ x_{2c} &= \left[\int p_c dt, p_c, \int q_c dt, q_c, \int r_c dt, r_c \right]^T \end{aligned} \quad (20)$$

$$\begin{aligned} x_1 &= \left[\int \alpha dt, \alpha, \int \beta dt, \beta, \int \mu dt, \mu \right]^T \\ \bar{x}_1 &= \left[\int \bar{\alpha} dt, \bar{\alpha}, \int \bar{\beta} dt, \bar{\beta}, \int \bar{\mu} dt, \bar{\mu} \right]^T \\ x_2 &= \left[\int p dt, p, \int q dt, q, \int r dt, r \right]^T \\ x_{2c} &= \left[\int p_c dt, p_c, \int q_c dt, q_c, \int r_c dt, r_c \right]^T \end{aligned} \quad (21)$$

以快回路为例, 快回路改写成如下形式:

$$\begin{aligned} \dot{x}_2 &= F_2(x_2) + G_2(x_2)u_2 \\ y_2 &= \omega \end{aligned} \quad (22)$$

其中:

$$\begin{aligned} F_1 &= [p, f_p, q, f_q, r, f_r]^T, u_2 = M_c \\ G_1 &= \begin{bmatrix} 0 & 0 & 0 \\ g_l^p & 0 & g_n^p \\ 0 & 0 & 0 \\ 0 & g_m^q & 0 \\ 0 & 0 & 0 \\ g_l^r & 0 & g_n^r \end{bmatrix} \end{aligned} \quad (23)$$

根据 TLC 方法设计思想, 定义增广系统状态误差 $e_2 = x_2 - x_{2c}$, 并将误差动态特性在 $(\bar{x}_2, \bar{\omega}_2)$ 处线性化, 可得

$$\begin{aligned} A_2(t) &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ a_{211}(t) & a_{212}(t) & a_{213}(t) & a_{214}(t) & a_{215}(t) & a_{216}(t) \\ 0 & 0 & 0 & 1 & 0 & 0 \\ a_{221}(t) & a_{222}(t) & a_{223}(t) & a_{224}(t) & a_{225}(t) & a_{226}(t) \\ 0 & 0 & 0 & 0 & 0 & 1 \\ a_{231}(t) & a_{232}(t) & a_{233}(t) & a_{234}(t) & a_{235}(t) & a_{236}(t) \end{bmatrix} \\ B_2 &= G_2 \end{aligned}$$

式中 $a_{2ij}(t)$ 具体表达式如下:

$$\begin{aligned} a_{211}(t) &= a_{213}(t) = a_{215}(t) = a_{221}(t) = a_{223}(t) \\ &= a_{225}(t) = a_{231}(t) = a_{233}(t) = a_{235}(t) = 0 \end{aligned}$$

$$\begin{aligned} a_{212}(t) &= I_{pq}^p \bar{q} + \bar{Q}S \frac{b^2}{2V} (g_l^p C_{l,p} + g_n^p C_{n,p}) \\ a_{214}(t) &= I_{pq}^p \bar{p} + I_{qr}^p \bar{r} \\ a_{216}(t) &= I_{qr}^p \bar{q} + \bar{Q}S \frac{b^2}{2V} (g_l^p C_{l,r} + g_n^p C_{n,r}) \\ a_{222}(t) &= 2I_{pp}^q \bar{p} + I_{pr}^q \bar{r} \\ a_{224}(t) &= g_m^q \bar{Q}SC_{m,q} \frac{c^2}{2V} \\ a_{226}(t) &= 2I_{rr}^q \bar{r} + I_{pr}^q \bar{p} \\ a_{232}(t) &= I_{pq}^r \bar{q} + \bar{Q}S \frac{b^2}{2V} (g_l^r C_{l,p} + g_n^r C_{n,p}) \\ a_{234}(t) &= I_{qr}^r \bar{r} + I_{pq}^r \bar{p} \\ a_{236}(t) &= I_{qr}^r \bar{q} + \bar{Q}S \frac{b^2}{2V} (g_l^r C_{l,r} + g_n^r C_{n,r}) \end{aligned}$$

取变换 $u_2(t) = T_2(t)v_2(t)$, 可将 B_2 变换成标准型:

$$T_2(t) = \begin{bmatrix} \frac{g_n^r}{g_l^p g_n^r - g_l^r g_n^p} & 0 & \frac{-g_n^p}{g_l^p g_n^r - g_l^r g_n^p} \\ 0 & \frac{1}{g_m^q} & 0 \\ \frac{-g_l^r}{g_l^p g_n^r - g_l^r g_n^p} & 0 & \frac{g_l^p}{g_l^p g_n^r - g_l^r g_n^p} \end{bmatrix} \quad (24)$$

定义期望的快回路闭环误差动态特性为

$$A_{2c2} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\tau_{211}(t) & -\tau_{212}(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\tau_{223}(t) & -\tau_{224}(t) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\tau_{235}(t) & -\tau_{236}(t) \end{bmatrix}$$

由

$$A_{2c2}(t) = A_{22}(t) + B_{22}(t)K_{22}(t) \quad (25)$$

可求得

$$K_{22}(t) = \frac{(A_{2c2}(t) - A_{22}(t))}{B_{22}(t)} \quad (26)$$

则系统真实的时变反馈增益矩阵

$$K_2(t) = T_2 \cdot K_{22}(t) \quad (27)$$

设 $K_1(t)$ 为慢回路系统的时变反馈增益矩阵, 慢回路控制器可用类似方法求得, 这里从略。

3.5 闭环 PD 谱设计

根据闭环系统 PD 谱理论可设计期望的闭环系统矩阵 A_{2c} 中的时变系数^[14-15]:

$$\begin{cases} \tau_{i11}(t) = \omega_{ni1}^2(t) \\ \tau_{i12}(t) = 2\zeta_{i1}\omega_{ni1}(t) - \frac{\dot{\omega}_{ni1}(t)}{\omega_{ni1}(t)} \end{cases} \quad (28)$$

$$\begin{cases} \tau_{i23}(t) = \omega_{ni2}^2(t) \\ \tau_{i24}(t) = 2\zeta_{i2}\omega_{ni2}(t) - \frac{\dot{\omega}_{ni2}(t)}{\omega_{ni2}(t)} \end{cases} \quad (29)$$

$$\begin{cases} \tau_{i35}(t) = \omega_{ni3}^2(t) \\ \tau_{i36}(t) = 2\zeta_{i3}\omega_{ni3}(t) - \frac{\dot{\omega}_{ni3}(t)}{\omega_{ni3}(t)} \end{cases} \quad (30)$$

其中 ζ_{ij} 为常值阻尼, $\omega_{nij}(t)$ 为时变带宽, 由于飞行控制系统采用时标分离原则对内外回路分别进行设计, 因此要求内回路的带宽相对于外回路带宽要足够大, 以便满足奇异摄动理论的要求^[15]。数值实验显示, 慢回路带宽取 2rad/s 较合适, 同时快回路带宽至少因为慢回路的 3 倍才能获得较好的性能。

4 仿真结果分析

4.1 控制器参数

采用数值方法研究控制系统的响应能力, 设计飞行器有 6 个控制面, 采用总偏转角最小的控制分配算法对控制力矩进行分配。系统实现时所需的控制器相关参数如下: 慢回路三通道的线性时变调节器阻尼均为 $\zeta_{1,i} = 0.7$, 时变带宽为 $\omega_{n1,j} = 2\text{rad/s}$, 伪微分带宽 $\omega_{1,diff} = 5\text{rad/s}$; 快回路的相应参数分别为 $\zeta_{2,i} = 0.7$, $\omega_{n2,j} = 10\text{rad/s}$, $\omega_{2,diff} = 15\text{rad/s}$ 。初始姿态角设为 $\alpha_0 = 0^\circ, \beta_0 = 0^\circ, \mu_0 = 0^\circ$, 初始角速度均为 0; 目标姿态角设为 $\alpha_c = 0^\circ, \beta_c = 0^\circ, \mu_c = 90^\circ$, 并通过低通滤波器 $1/(1+s)$ 滤波。

4.2 40km 高度仿真

首先选取 40km 高度, 马赫数 15, 仿真结果如图 8 所示。此时控制器能够实现对飞行器的姿态控制, 各个控制面偏转角均小于 10° 。

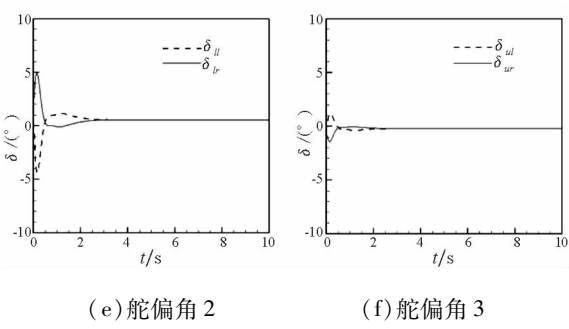
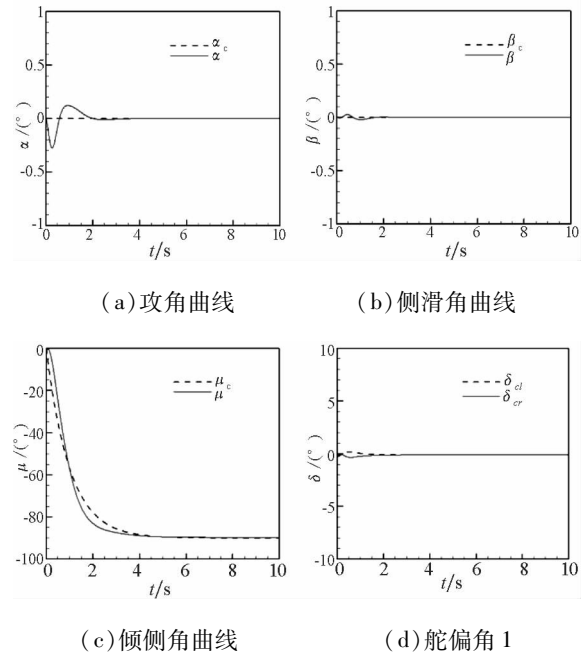


图 8 40km 高度仿真结果

Fig. 8 Simulation results of 40km

4.3 60km 高度仿真

高度增加到 60km 时, 马赫数 15 条件下, 直接按上述方法给制导指令, 则控制系统将发散, 原因如前所述, 控制系统根据制导指令计算得到的控制力矩超过控制面的控制能力, 改变控制策略, 制导指令采用分段的方法:

$$\mu_c = \begin{cases} -30^\circ, & t < 1\text{s} \\ -60^\circ, & 1\text{s} \leq t < 2\text{s} \\ -90^\circ, & t \geq 2\text{s} \end{cases} \quad (31)$$

仿真结果如图 9 所示。控制面偏转角度和 40km 高度相比有较大的提高, 这主要是由于 60km 高空空气密度远小于 40km 高空的空气密度。

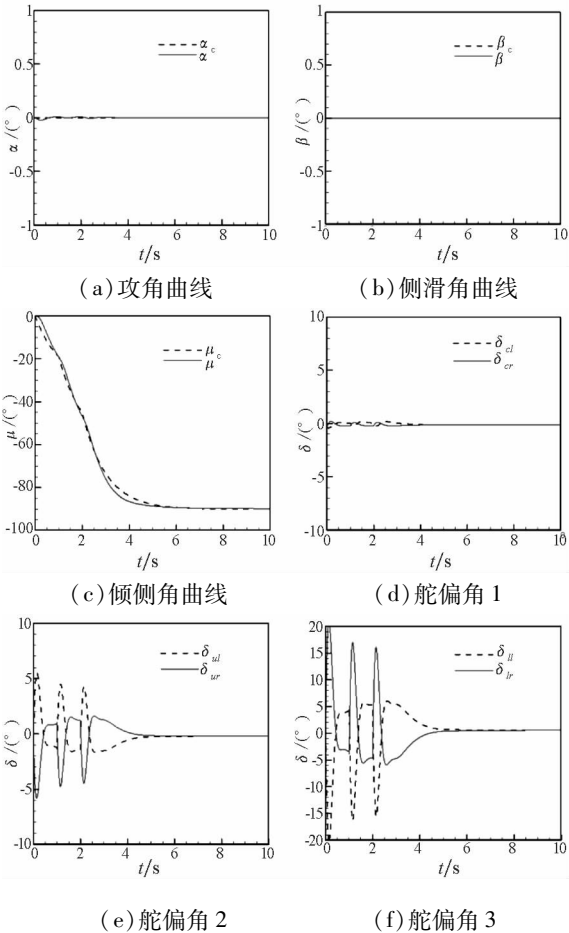


图 9 60km 高度仿真结果

Fig. 9 Simulation results of 60km

可以看出,在高空条件下,进行轨迹优化及控制时,应结合控制系统的实现能力,给定相关约束(主要是滚转角速度约束),以完成设计目标,验证了第二部分的分析结果。

仿真过程中,控制器均取相同参数,这表明,所设计的飞行器具有较强的鲁棒性。

5 结 论

本文针对与高超声速滑翔飞行器机动性能相关的倾斜转弯技术进行了研究。以平衡滑翔弹道为参考弹道,研究了倾斜转弯转弯半径、下降高度以及倾侧角之间的关系,提出在对控制系统进行设计时,应根据高度的差别考虑执行机构的响应能力,合理设计控制指令。

应用轨迹线性化方法设计了飞行器倾斜转弯姿态控制系统。根据奇异摄动理论将动力学系统分为快慢回路两部分,针对快慢回路分别设计控制器。对控制系统的性能进行了仿真分析,从仿真结果可以看出,设计的控制器具有良好的控制性能,但随着高度的增加,控制指令应结合实际控制能力,以完成对飞行器的姿态控制。

参考文献 (References)

[1] 李菁菁, 任章, 黎科峰. 高超声速飞行器再入段的动力学建模与仿真 [J]. 系统仿真学报, 2009, 21(2): 534 – 537.
LI Jingjing, REN Zhang, LI Kefeng. Flight dynamics modeling and simulation of hypersonic vehicles in reentry phase [J]. Journal of System Simulation, 2009, 21(2): 534 – 537. (in Chinese)

[2] 李扬, 陈万春. 高超声速飞行器 BTT 非线性控制器设计与仿真 [J]. 北京航空航天大学学报, 2006, 32(3): 249 – 253.
LI Yang, CHEN Wanchun. Design and simulation of BTT nonlinear controller for hypersonic vehicle [J]. Journal of Beijing University of Aeronautics and Astronautics, 2006, 32(3): 249 – 253. (in Chinese)

[3] 钱承山, 吴庆宪, 姜长生. 空天飞行器概念设计再入数学建模研究 [J]. 宇航学报, 2008, 29(2): 435 – 441.
QIAN Chengshan, WU Qingxian, JIANG Changsheng.

Atmospheric reentry modeling of aerospace vehicle dynamics for conceptual design Studies [J]. Journal of Astronautics, 2008, 29(2): 435 – 441. (in Chinese)

[4] Wang Y H, Wu Q X, Jiang C S, et al. Guaranteed cost fuzzy output feedback control via LMI method for re-entry attitude dynamics. [J]. Journal of Uncertain Systems. 2007, 21(1): 291 – 302.

[5] 黄国勇, 姜长生, 王玉惠. 基于快速模糊干扰观测器的 UASV 再入 Terminal 滑模控制 [J]. 宇航学报, 2007, 28(2): 292 – 297.
HUANG Guoyong, JIANG Changsheng, WANG Yuhui. Research of terminal sliding mode control based on fast fuzzy disturbance observer for UASV ReEntry. [J]. Journal of Astronautics, 2007, 28(2): 292 – 297. (in Chinese)

[6] 张春雨, 姜长生, 朱亮. 基于模糊干扰观测器的空天飞行器轨迹线性化控制 [J]. 宇航学报, 2007, 28(1): 33 – 38.
ZHANG Chunyu, JIANG Changsheng, ZHU Liang. Trajectory linearization control for an aerospace vehicle based on fuzzy disturbance observer [J]. Journal of Astronautics, 2007, 28(1): 33 – 38. (in Chinese)

[7] Snell S A. Nonlinear Dynamic-Inversion flight control of supermaneuverable aircraft [D]. University of Minnesota. 1991.

[8] Mooij I E. The motion of a vehicle in a planetary atmosphere [M]. Delft, The Netherlands: Delft University Press, 1997.

[9] Da Costa R R. Reentry attitude controller development using nonlinear dynamic inversion for the crew return vehicle [D]. Delft: Delft University of Technology. 2004.

[10] Vinh N X. Hypersonic and planetary entry flight mechanics [M]. Michigan: The University of Michigan Press, 1980.

[11] De OlivC F L. Nonlinear dynamics and stability of hypersonic reentry vehicles [D]. Michigan: University of Michigan, 1995.

[12] Bevacqua T, Best E, Huizenga A, et al. Improved trajectory linearization flight controller for reusable launch vehicles [R]. AIAA 2004 – 875. 2004.

[13] Chen X Q, Hou Z X, Liu J X. Nonlinear bank-to-turn autopilot design for hypersonic gliding vehicle [C]// International Symposium on Systems and Control in Aeronautics and Astronautics 2010. Harbing, China, 2010.

[14] Zhu J J. Nonlinear tracking and decoupling by trajectory linearization [R]. 1998.

[15] Zhu J J, Huizenga A B. A type two trajectory linearization controller for a reusable launch vehicle-a singular perturbation approach [R]. AIAA 2004 – 5184, 2004.

遥操作交会对接系统研究*

周剑勇, 张 波, 蒋自成, 李海阳
(国防科技大学 航天与材料工程学院, 湖南 长沙 410073)

摘 要: 遥操作交会对接是指操作人员在远端通过遥操作方式控制追踪航天器进行交会对接, 主要用于无人航天器自动交会对接系统故障条件下平移靠拢段的交会对接控制。简要介绍了遥操作交会对接的基本概念及其国内外研究现状; 结合我国载人航天工程的基本特点, 设计了适用于我国未来空间站任务的遥操作交会对接系统; 根据设计的系统方案, 设计和开发了相应的遥操作交会对接仿真系统, 并进行了部分仿真分析和遥操作交会对接试验, 初步验证了设计方案的可行性。

关键词: 遥操作交会对接; 系统设计; 仿真分析; 时延
中图分类号: V526 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0024-05

Researches on teleoperation rendezvous and docking system

ZHOU Jianyong, ZHANG Bo, JIANG Zicheng, LI Haiyang

(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Teleoperation rendezvous and docking is referred to as the concept that the chase spacecraft is flown distantly through teleoperation approach. It can be used as a backup for autonomous rendezvous and docking for the unmanned chases during the finally approach stage. The concept of teleoperation rendezvous and docking (RVD) and the previous studies in this field were introduced firstly. Then, a design of the teleoperation RVD system was addressed, which is based on the primary assumed conditions of our future space station project. Finally, a simulation system was developed and corresponding experiments were carried out for verifying the design, especially the design with teleoperation time delay. The results show that the design is feasible, even though there are jobs to be done in the future.

Key words: teleoperation rendezvous and docking; system design; simulation analysis; time delay

随着我国载人航天事业的不断发展以及空间站工程的论证、立项等工作的开展,我国将开展载人空间站的建造。在空间站的建造、后续运营及功能拓展中,需要进行无人航天器与空间站的大量交会对接活动,如货运飞船、实验舱等与空间站的交会对接。由于无人航天器不载人的特性,无法采用载人飞船的手动交会对接方式作为自动交会对接的备份,一旦自动交会对接系统出现故障,将不得不放弃整个交会对接任务,由此将造成巨大的经济损失和社会影响,甚至个别情况下还将威胁到空间站的安全。

因此,研究一种无人航天器自动交会对接系统的备份手段,用于进行自动系统异常情况下的交会对接控制,具有重要的工程实践价值、经济效益和社会意义。

1 遥操作交会对接概述

1.1 遥操作交会对接概念

遥操作交会对接是指操作人员在远端(或不在控制飞行器环境内)通过遥操作方式操作追踪航天器,实现交会对接,国外有学者称之为交会对接的遥操作控制(teleoperation control)^[1],也有学者称之为远程控制(remote control)^[2]、遥操作飞行(remote pilot、teleoperation pilot)^[3-4]等。就概念本身而言,遥操作交会对接借鉴了空间机器人遥操作的概念,是传统遥操作概念的一个延伸,是遥科学的一个子领域。

国内外有相关学者和研究人员曾提出过“遥操作航天器”(teleoperator spacecraft)的概念,并提出了遥操作航天器的总体设计思路等^[5-7]。遥操作航天器是指可以通过地面操作人员或空间站上航天员进行遥操作控制的一类航天器,用于进

* 收稿日期:2011-09-06
基金项目:国家部委资助项目
作者简介:周剑勇(1983—),男,江西宜春人,博士研究生,E-mail:zhoujianyong@nudt.edu.cn;
蒋自成(通信作者),男,湖南常德人,讲师,博士,E-mail:jiangzicheng@nudt.edu.cn

行交会对接、在轨服务、目标捕获等空间活动。

遥操作航天器任务实施过程中采用的空间交会对接技术就是遥操作交会对接技术,相关研究文献简要阐述了遥操作航天器的概念和基本设计思想。

遥操作交会对接主要有两种工作模式,一是空间站上航天员对追踪航天器进行遥操作交会对接控制的工作模式;一是地面操作人员对追踪航天器进行遥操作交会对接控制的工作模式,可分别简称为站-船遥操作交会对接模式和地-船遥操作交会对接模式。遥操作交会对接操作过程中,操作员主要利用反馈的 TV 摄像机图像及相对运动位置、速度等测量信息,通过遥操作平台在线实时地控制远端追踪航天器完成交会对接。

1.2 遥操作交会对接研究现状

美国曾于 20 世纪 80 年代在 OMV (Orbital Maneuvering Vehicle) 项目中开展过遥操作交会对接技术的基础研究^[8]。项目计划研制一种通用的遥操作航天器,用于进行在轨服务、回收废弃卫星等空间活动,后来由于研制经费等因素,项目于 1991 年下马。

俄罗斯(前苏联)则于 20 世纪 90 年代初,研制成功了站-船模式的 TORU (Teleoperatormiy Rezhim Upravleniya) 遥操作交会对接系统,并应用于和平号(Mir)空间站和国际空间站(ISS)的交会对接任务,作为无人货运飞船自动交会对接系统的备份。近年来,TORU 系统已成功实施多次进步号无人货运飞船自动交会对接系统故障下与 ISS 的交会对接^[9],取得了巨大的经济效益。

日本 NASDA (National Space Development Agency of Japan)在 ETS(Engineering Test Satellite)计划中也开展了相关研究,并计划在 ETS 7 任务

期间进行基于地面的地-船模式遥操作交会对接空间实验,最终由于种种原因,实验没有如期进行^[3-4,10-11]。

相对于其他航天大国或主要的空间技术研究组织,从国内学者公开发表的研究文献来看,国内航天领域的各研究机构、科研院所等对遥操作交会对接的研究还未正式开展。

2 遥操作交会对接系统设计

综合 TORU 系统和 ETS 7 的遥操作交会对接系统等,得知国外研究的遥操作交会对接系统往往只支持站-船模式或地-船模式单一的遥操作模式。这与各国的航天发展基本国情有一定的关联,俄罗斯由于基本解决了空间站航天员长期驻守的相关技术,且其参与建设与维护的空间站(包括 ISS 和 Mir)有航天员长期驻守,因此只需要研究和实现站-船模式的遥操作交会对接即可满足其空间站的应用需求。相对而言,日本一直希望独立地开展自己的航天活动,并且 ISS 上已有 TORU 系统可实现站-船模式的遥操作交会对接,故 NASDA 重点转向地-船模式的遥操作交会对接研究。结合我国的载人航天现状,我国未来空间站将面临无航天员值守的情况。因此,我们的遥操作交会对接系统需要同时支持站-船模式和地-船模式两类遥操作交会对接的工作模式。当自动交会对接系统出现故障时,根据实际情况启动遥操作交会对接系统,由航天员进行站-船模式的遥操作交会对接控制或由地面操作人员进行地-船模式的遥操作交会对接控制。

根据以上的总体任务需求,初步设计我国未来空间站的遥操作交会对接系统,如图 1 所示。

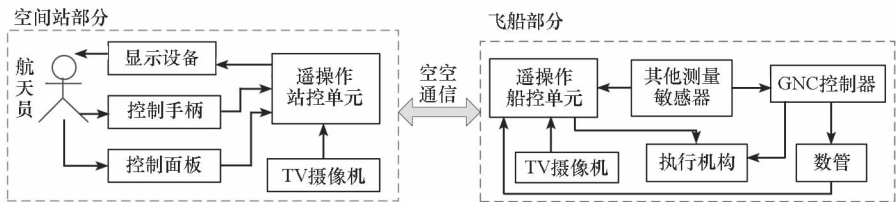


图 1 站-船遥操作交会对接系统原理

Fig. 1 Space station based teleoperation RVD

图 1 中,遥操作站控单元和遥操作船控单元为遥操作交会对接的运动控制单元。航天员根据安装在飞船上交会对接摄像机、安装在空间站上的舱外监视摄像机等的电视图像,结合其他敏感器的测量信息(在显示设备上显示),遥操作货运飞船等无人追踪器进行交会对接或撤离,测量信息和遥操作指令通过空空通信或天地通信系

统进行传输。控制手柄主要进行飞船平移运动和姿态运动的遥操作控制,控制面板主要用于施加部分其他控制指令,如施加避撞机动(CAM)等。

地-船遥操作交会对接的控制流程与站-船模式部分类似,遥操作环境由地面遥操作系统组成,如图 2 所示。相对于站-船模式空间站的操作环境,地面环境可利用的资源显著增多,如更强

的计算能力、更先进的人机交互设备等,设计的地-船遥操作交会对接系统充分利用地面丰富的资源,引入各种先进的遥操作技术,如预测控制技术、预显示技术、虚拟现实技术等,以克服和解决天地大时延条件下的遥操作交会对接问题。此外,地-船遥操作模式下,个别极端情况下,地面操作员可能需要遥操作空间站进行紧急避撞机动,以确保空间站的安全。

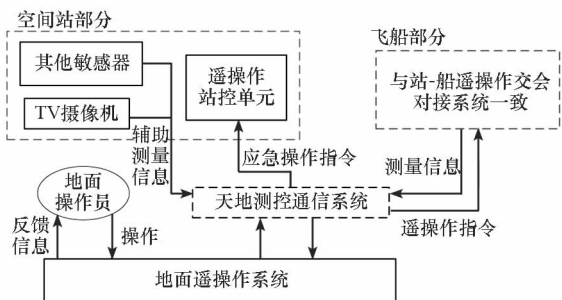


图2 地-船遥操作交会对接系统原理图

Fig. 2 Ground based teleoperation RVD

分析以上设计方案,遥操作交会对接与载人飞船的手动交会对接部分类似,操作者的主要信息来源和操作依据为TV图像及其他测量设备的辅助测量信息,而TV图像的传输可能需要较大的通信带宽,这是遥操作交会对接系统的一个重要约束。

此外,进行地-船遥操作交会对接时,遥操作时延可达 $5 \sim 7\text{s}^{[12-13]}$,大时延对遥操作交会对接的操作效率、成功率、安全性等均有影响,需要深入分析和研究时延对遥操作交会对接控制、安全性等的影响特性。

3 仿真系统设计与仿真分析

针对本文设计的遥操作交会对接系统,设计并研制相应的遥操作交会对接仿真系统,进行设计方案的初步分析和仿真验证。尤其对设计方案中两个重要的约束进行分析和仿真试验,即分析TV图像的传输带宽是否可满足工程约束、操作时延的影响情况或时延影响是否可克服。

3.1 仿真系统设计

(1) 基本仿真模型

飞船和空间站的动力学模型采用高精度的轨道积分模型,包括J2项摄动、大气阻力摄动、第三体引力摄动等摄动模型^[14]。空间站处于姿态稳定的被动飞行模式,不配合飞船进行交会对接机动;飞船的姿态由GNC系统进行自动控制,遥操作只进行飞船的平移运动控制。

(2) 仿真系统总体框架

仿真系统主要进行飞船/空间站的动力学与控制的仿真、TV图像的模拟、图像的编码与图像显示等。仿真系统的组成和总体信息流关系如图3所示。

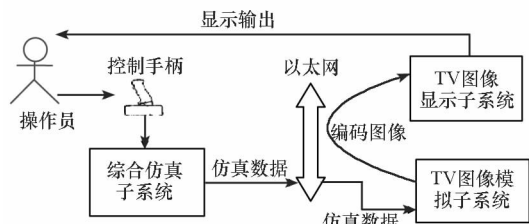


图3 仿真系统总体结构

Fig. 3 The teleoperation RVD simulation system

综合仿真子系统主要实现对仿真系统的综合管理;进行激光雷达等测量敏感器的功能仿真;进行空间站GNC系统功能的仿真,包括轨道运动的自主控制和对地定向姿态运动的自主控制;进行飞船GNC系统部分功能的仿真,主要为姿态运动的自动控制和飞船平移运动的动力学仿真。

TV图像模拟子系统通过三维可视化的方式模拟TV摄像机的场景图像,并对TV图像的压缩编码过程进行模拟,将编码压缩后的图像信息通过网络发送至TV图像显示子系统,进行图像显示。

TV图像显示子系统通过网络接收TV摄像机图像模拟子系统的图像数据,并对图像数据进行解码得到图像信息,将解码后的图像信息显示在显示终端设备,模拟显示终端的TV图像显示。

TV图像的模拟中引入了类工程的图像编码压缩、数据传输和解压显示的过程,既提高了图像模拟的真实性,又可用于进行图像传输带宽的测试以及分析不同图像质量对遥操作交会对接的影响等。

3.2 仿真试验及结果分析

仿真试验主要对前面提到的两个系统设计约束进行操作试验和分析,试验结果分析如下。

(1) 图像通信带宽分析

本仿真系统中TV图像的压缩格式采用JPEG格式,逐帧对图像进行压缩和传输,图像为24位真彩色模式,综合试验结果分析如表1所示。

分析上图,25f/s的连续图像,保证图像较好的清晰度条件下,通信带宽的需求约 $3.8 \sim 4.8\text{Mb/s}$ 。

表 1 图像通信带宽测试

Tab. 1 Communication bandwidth experimental results

分辨率	帧率(f/s)	图像质量	平均带宽 (Mb/s)
1024 × 768	25	一般	3.9
		较清晰	4.8
		很清晰	6.1
720 × 576	25	一般	2.7
		较清晰	3.8
		很清晰	5.9

参考 ATV^[15]、ETS 7^[16] 等遥操作系统的设计,其通信带宽约 1.5Mb/s。实际工程中,图像为 8 位灰度图像,且压缩算法为经过优化的 MPEG 流压缩算法时,通信带宽将进一步显著减小,可降至 1.5Mb/s 左右。因此,TV 图像的传输带宽可满足工程约束,以 TV 图像为遥操作的主要参考信息是可行的。

(2)遥操作时延影响分析

由于地-船遥操作交会对接模式的操作时延为 5~7s,参考国内外对空间遥操作机器人仿真时延的设定^[17-19],仿真系统的闭环操作时延设为 6s,单向时延 3s,时延的模拟通过软件算法实现。

试验初始条件设为飞船在空间站正后方 150m,采用 -V Bar 方式进行交会对接。实验操作人员直接面向延迟后的图像和测量模拟数据进行操作,暂未采用预显示技术等辅助操作。试验任务时间要求不大于 720s,以确保任务操作均在地面可监视弧段内。遥操作交会对接试验成功与否通过比较终端状态是否满足对接要求^[20],对接终端要求定义如表 2 所示。

表 2 对接精度要求

Tab. 2 Expected final docking states

项目	要求
接近速度	$0.05\text{m/s} \leq V_x \leq 0.4\text{m/s}$
横向速度	$\sqrt{V_y^2 + V_z^2} \leq 0.3\text{m/s}$
横向位移	$\sqrt{y^2 + z^2} \leq 0.3\text{m}$

表 3 中,0s 时延与载人飞船的手动交会对接类似,认为闭环控制回路无显著时延。2s 时延接近于站-船遥操作交会对接的闭环时延,其结果可作为站-船模式的参考试验结果。6s 时延与地-船模式的时延接近,反映地-船模式大时延的操作试验情况。

表 3 时延影响试验

Tab. 3 Results of time delay experiment

时延条件	燃料消耗 均值(kg)	任务时间 均值(s)	失败次数	成功次数
0s	9.05	578.52	0	20
2s	9.31	591.36	0	20
6s	9.71	619.35	4	16

分析以上试验结果,遥操作大时延对遥操作交会对接的成功率影响显著,随着操作时延的增大,遥操作交会对接任务的燃料消耗逐渐增加,任务操作时间也逐渐增长。

4 结 论

根据以上的系统设计和试验分析,可知:

(1)提出的遥操作交会对接系统设计结合了我国载人航天的实际情况,系统资源需求可满足工程约束,设计方案具有一定的合理性和可行性。

(2)地-船模式的遥操作大时延对遥操作交会对接影响显著,需要进一步研究克服遥操作时延的相关技术。

(3)遥操作交会对接不仅可用于我国未来空间站任务中,还可拓展应用于非合作目标的空间交会与捕获等,具有重要的工程应用价值和潜在的军事价值,应就此技术展开广泛而深入的研究。

参考文献(References)

[1] Jones S. Orbital maneuvering vehicle teleoperation and video data compression [J]. AIAA Journal, 1989, 28(9): 99-117.

[2] Cislighi M. The ATV rendezvous pre-development program (ARP) [C]// Proceedings of the 22nd AAS Guidance Control Conference. Breckenridge; ESA 1999.

[3] Mitsushige O. Experiences and lessons learned from the ETS-VII robot satellite [C]// Proceedings of IEEE Internal Conference on Robotics and Autonomous. San Francisco; IEEE, 2000: 914-919.

[4] Masaaki M, Isao K. Development of ETS-VIII RVD system-preliminary design and EM development phase engineering test satellite and rendezvous docking [C]// Proceedings of the AIAA Guidance, Navigation, and Control Conference. Washington DC; AIAA, 1995. 1656-1664.

[5] Tewell J R, Richard A S. Advanced teleoperator spacecraft [C]// Proceedings of the AIAA Conference on Large Space Platforms: Future Needs and Capabilities. Los Angeles; AIAA, 1978.

[6] Tuner J R, Spencer R A. Teleoperator Maneuvering System [C]// Proceedings of the AIAA Conference on Space Systems. Washington DC; AIAA, 1982.

[7] Zhou J Y, Jiang Z C, Tang G J. Conceive of the conception

- and investigation of teleoperator spacecraft [C]// 61st International Astronautical Congress, Prague; International Astronautical Federation 2010.
- [8] Keith E C. OMV mission simulator[C]// Proceedings of the IEEE Conference on Space Technology. Anaheim; AIAA, 1990.
- [9] Zhou J Y, Jiang Z C, Tang G J. A new approach for teleoperation rendezvous and docking with time delay[J]. Sci China-Phys Mech Astron, 2012, 55: 1–6.
- [10] Mitsushige O, Isao K. ETS-7, a rendezvous docking and space robot technology experiment satellite result of the engineering model development work [C]// Proceedings of the 34th SICE Annual Conference. Hokkaido; IEEE, 1995.
- [11] Isao K, Masaaki M, et al. In-orbit demonstration of an unmanned automatic rendezvous and docking system by the Japanese engineering test satellite ETS-VII[C]// Proceedings of the AIAA Guidance, Navigation, and Control Conference. Washington; AIAA, 1994; 950–960.
- [12] Sheridan T B. Space teleoperation through time delay: review and prognosis [J]. IEEE Transactions on Robotics and Automation, 1993, 9(6): 592–606.
- [13] Noriaki A, Lee J H, et al. A study on influence of time delay in teleoperation[C]// Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics. Atlanta; IEEE, 1999; 317–322.
- [14] 王华, 唐国金, 李海阳. 航天系统分析与仿真基础程序库 AstroLib[J]. 系统仿真学报, 2007, 19(13): 2917–2920.
- WANG Hua, TANG Guojin, LI Haiyang. AstroLib: a foundational library for aerospace system analysis and simulation[J]. Journal of System Simulation, 2007, 19(13): 2917–2920. (in Chinese)
- [15] Gianni C, Giovanni B P. Visual techniques applied to the ATV-ISS rendezvous monitoring [C]// Proceedings of the IEEE Aerospace Conference. Atlanta; IEEE, 2004; 613–625.
- [16] Mitsushige O. System engineering approach in designing the teleoperation system of the ETS-VII robot experiment satellite[C] // Proceedings of the IEEE International Conference on Robotics and Automation. New Mexico; IEEE, 1997; 3050–3061.
- [17] Yoon W K, Toshihiko G, Hiroshi K. Model based space robot teleoperation of ETS-VII manipulator[J]. IEEE Transactions on Robotics and Automation, 2004, 20(3): 602–612.
- [18] 李成, 梁斌. 空间机器人的遥操作[J]. 宇航学报, 2001, 22(1): 95–98.
- LI Cheng, LIANG Bin. Teleoperation of space robotics[J]. Journal of Astronautics, 2001, 22(1): 95–98. (in Chinese)
- [19] 曾庆军, 宋爱国, 黄惟一. 时延下空间遥操作机器人系统工作模式研究[J]. 宇航学报, 2003, 24(2): 180–184.
- ZENG Qingjun, SONG Aiguo, HUANG Weiyi. Research on operation mode of space teleoperation robot with time delay[J]. Journal of Astronautics, 2003, 24(2): 180–184. (in Chinese)
- [20] 张波, 蒋自成, 周剑勇, 等. 交会对接平移靠拢段手控策略设计[J]. 载人航天, 2010, 16(2): 21–25.
- ZHANG Bo, JIANG Zicheng, ZHOU Jianyong, et al. The hand control strategy of rendezvous and docking in the approach phase[J]. Manned Spaceflight. 2010, 16(2): 21–25. (in Chinese)

临近空间飞行器纵向逆控制系统设计*

王 鹏,刘鲁华,吴 杰

(国防科技大学 航天与材料工程学院,湖南 长沙 410073)

摘 要:建立了临近空间飞行器无动力滑翔阶段的纵向运动模型,并应用动态逆方法推导了模型的逆系统。系统相对阶等于绕质心运动模型的系统阶数,实现了绕质心运动模型的线性化。为了克服纵向运动模型的隐动态和参数的不确定性,在内环动态逆控制器的基础上,设计了外环最优调节器。仿真结果表明:飞行器纵向逆控制系统可以准确跟踪俯仰角或攻角指令,并对参数不确定性和外界干扰具有较好的鲁棒性,具有工程适用性。

关键词: 临近空间;滑翔;纵向模型;动态逆;最优控制;鲁棒性

中图分类号: V412.4 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0029-04

Longitudinal inversion control system design
for near-space vehicle

WANG Peng, LIU Luhua, WU Jie

(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Near-space glide vehicle motion was modeled by nonlinear longitudinal dynamic equations. The Inverse system of this model was obtained by using dynamic inversion approach. The relative degree of dynamic model was equal to the rotational motion system order, so rotational motion dynamic model could be Input/Output linearization completely. The dynamic inversion method and optimal control were used in inner-loop and outer-loop for the longitudinal dynamic model. Simulation studies demonstrate that the proposed controller tracks the pitch angle or angle of attack command accurately, meets the performance requirements with acceptable control inputs and is robust with respect to the parametric uncertainties and external disturbance.

Key words: near-space; glide; longitudinal model; dynamic inversion; optimal control; robustness

滑翔飞行器是一类临近空间飞行器,它利用固体火箭发动机助推,到达一定高度和速度后,助推器关机并与飞行器分离,之后飞行器以高超声速无动力滑翔再入。由于这类飞行器飞行空域、速度的跨度和变化都非常大,体现出快时变、强耦合、强非线性和强不确定的特点,因此对这类对象的控制,传统增益预置的线性控制方法难以达到满意的控制效果,还会带来分段过多、控制器频繁切换的问题,而非线性控制方法则提供了更好的解决方案^[1-7]。

非线性动态逆控制又称为反馈线性化方法,是20世纪90年代,Senll、Enns和Garrard等人针对飞机在做大迎角超机动飞行时的本体非线性、强耦合特征而提出来的非线性控制策略^[1],其基本思想是:对于具体的研究对象,用系统模型生成

一种可用反馈方法实现的原系统的“ α 阶积分逆系统”,将对象补偿为具有线性传递关系的伪线性系统。本文利用动态逆方法处理非线性问题,在分析模型可逆性的基础上,设计了基于最优控制的临近空间飞行器纵向逆控制系统,并通过仿真计算分析了控制系统对参数不确定性和外界干扰的鲁棒性。

1 临近空间飞行器纵向模型

为了推导临近空间飞行器无动力滑翔阶段的纵向模型,提出如下假设:视地球为均质圆球,并忽略地球旋转;飞行器为面对称体,惯量积 $J_{xz} = J_{yz} = 0$,惯量积 J_{xy} 为小量,忽略不计;侧向运动参数 $z, \beta, \sigma, \gamma_v, \psi, \gamma, \omega_x, \omega_y$ 均为小量。基于上述假设,得到飞行器标量形式的纵向运动模型为

* 收稿日期:2011-09-18

基金项目:国家部委资助项目

作者简介:王鹏(1984—),男,吉林洮南人,博士研究生,E-mail:wonderful2020@163.com;

吴杰(通信作者),男,教授,博士,博士生导师,E-mail:wujie_nudt@sina.com

$$\begin{cases} \dot{V} = -\frac{\mu}{r^3} [x \cos \theta + (y + R_e) \sin \theta] - \frac{D}{m} \\ \dot{\theta} = -\frac{\mu}{r^3 V} [-x \sin \theta + (y + R_e) \cos \theta] + \frac{L}{mV} \\ \dot{\omega}_z = \frac{M_z}{J_z}; \dot{\varphi} = \omega_z; \dot{x} = V \cos \theta; \dot{y} = V \sin \theta \end{cases} \quad (1)$$

其中, V 为飞行器速度; θ 为速度倾角; ω_z 为俯仰角速度; x, y 为飞行器在地面坐标系中的位置坐标; φ 为俯仰角; $\mu = 3.98600436 \times 10^{14} \text{ (m}^3/\text{s}^2\text{)}$ 为地球引力常数; $r = \sqrt{x^2 + (y + R_e)^2}$ 为地心距, R_e 为球形引力场时的地球平均半径; D, L 分别为阻力和升力; M_z 为俯仰力矩; m 为飞行器质量; J_z 为飞行器 z 轴转动惯量。

攻角 α 与俯仰角、速度倾角之间满足如下关系

$$\alpha = \varphi - \theta \quad (2)$$

阻力、升力和俯仰力矩的表达式分别为

$$\begin{cases} D = qSC_D \\ L = qSC_L \\ M_z = qSlm_z \end{cases} \quad (3)$$

其中, $q = \frac{1}{2} \rho V^2$ 为动压, ρ 为空气密度; S 为飞行器参考面积; l 为飞行器特征长度; C_D, C_L, m_z 分别为阻力系数、升力系数和俯仰力矩系数。

本文所研究的飞行器采用气动舵单一控制方式, 气动力系数和力矩系数为飞行马赫数 Ma 、高度 H 、攻角 α 及俯仰舵偏角 δ_φ 的非线性函数, 可表达为如下关系式

$$\begin{cases} C_D = f_D(Ma, H, \alpha, \delta_\varphi) \\ C_L = f_L(Ma, H, \alpha, \delta_\varphi) \\ m_z = f_z(Ma, H, \alpha, \delta_\varphi) \end{cases} \quad (4)$$

2 飞行器纵向运动模型的可逆性分析

以飞行器速度 V 、速度倾角 θ 、俯仰角速度 ω_z 、地面坐标系中的质心坐标 x, y , 以及俯仰角 φ 作为飞行器纵向动力学模型的状态变量, 即

$$\mathbf{x} = [V \quad \theta \quad \omega_z \quad x \quad y \quad \varphi]^T \quad (5)$$

且输入量取为俯仰舵偏角, 即 $u = \delta_\varphi$, 输出量取为飞行器俯仰角, 即 $y = \varphi$, 则飞行器纵向运动模型可表述为如下的状态方程形式

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{f}_u(\mathbf{x}, u) \\ y = \varphi \end{cases} \quad (6)$$

其中

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} -\frac{\mu}{r^3} [x \cos \theta + (y + R_e) \sin \theta] \\ -\frac{\mu}{r^3 V} [-x \sin \theta + (y + R_e) \cos \theta] \\ 0 \\ V \cos \theta \\ V \sin \theta \\ \omega_z \end{bmatrix}$$

$$\mathbf{f}_u(\mathbf{x}, u) = qS \begin{bmatrix} -\frac{C_D}{m} & \frac{C_L}{mV} & \frac{l}{J_z} m_z & 0 & 0 & 0 \end{bmatrix}^T$$

可见纵向动力学模型是一个 SISO (单输入单输出) 系统。为了分析飞行器运动模型的可逆性, 首先给出多变量频域理论中 MIMO 非线性系统的可逆性定理——函数可控性定理。

定理 1^[8] 一般 MIMO 非线性系统在 $(\mathbf{x}_0, \mathbf{u}_0)$ 的邻域内可逆的充分必要条件是在此邻域内存在相对向量阶, 即 $(\mathbf{x}_0, \mathbf{u}_0)$ 为 Interactor 算法的正则点。

SISO 系统是 MIMO 系统的特殊情况, 故 (6) 式所示的 SISO 系统在 $(\mathbf{x}_0, \mathbf{u}_0)$ 的某个邻域内可逆的充分必要条件是在此邻域内存在相对阶 k 。求解系统相对阶的基本方法是 Interactor 算法^[8]。该算法的基本思想是: 输出的每个分量分别对时间求导数, 直到导数中显含输入量为止, 即

$$\frac{\partial y_j^{(r_j)}}{\partial u} \neq 0 (j = 1, 2, \dots, m) \quad (7)$$

m 为输出的维数, 此时所求导数的次数 r_j 即为该分量的相对阶。

下面利用 Interactor 算法分析飞行器纵向动力学模型系统的可逆性。首先, 计算飞行器俯仰角 φ 关于时间 t 的导数, 直到导数表达式中显含输入量为止。经推导得

$$\dot{\varphi} = \omega_z; \quad \ddot{\varphi} = \dot{\omega}_z = \frac{M_z}{J_z} \quad (8)$$

令 $\mathbf{Y} = \ddot{\varphi}$, 则有

$$\frac{\partial \mathbf{Y}}{\partial u} = \frac{M_z^{\delta_\varphi}}{J_z} = \frac{\rho V^2 S l}{2 J_z} m_z^{\delta_\varphi} \quad (9)$$

显然有 $\frac{\partial \mathbf{Y}}{\partial u} \neq 0$, 至此 Interactor 算法正常结束, 相对阶 $k = 2$, 满足 $k < n = 6$, 故飞行器纵向运动模型是可逆的。

由以上可逆性推导过程可知: 可逆性分析只用了如下的绕质心运动模型

$$\begin{cases} \dot{\omega}_z = \frac{M_z}{J_z} \\ \dot{\varphi} = \omega_z \end{cases} \quad (10)$$

并有系统相对阶等于绕质心运动模型的系统阶数。当系统相对阶等于系统阶数时,利用反馈线性化方法可实现模型的完全线性化^[8],故动态逆方法实现了绕质心运动模型的线性化。

若记绕质心运动的原系统 Σ 为

$$\varphi = \Phi(\delta_\varphi)$$

(11)

原系统以俯仰舵偏角 δ_φ 为输入,俯仰角 φ 为输出,则可构造逆系统 $\widehat{\Sigma}$ 为

$$\delta_\varphi = \Xi(\ddot{\varphi})$$

(12)

原系统 Σ 与逆系统 $\widehat{\Sigma}$ 串联,可得到如下的伪线性复合系统

$$\varphi = \Phi[\Xi(\ddot{\varphi})] = \Gamma(\ddot{\varphi})$$

(13)

可见引入逆系统后,伪线性复合系统的输入输出之间具有线性传递关系,理论上可将该线性关系简单地表示成如下的线性积分型传递函数

$$G(s) \approx \frac{1}{s^2}$$

(14)

但这种线性关系仅限于输入输出之间,系统的内部结构可能仍是非线性关系,这也就是伪线性的含义。

3 基于最优控制的纵向逆控制系统设计

虽然动态逆方法实现了绕质心运动模型的线性化,但由于系统相对阶小于纵向运动模型的总系统阶数,故纵向运动存在隐动态,内环动态逆控制器达不到期望的性能指标。从充分利用绕质心运动线性化结果和解决隐动态控制问题的角度出发,可采用线性系统理论设计外环最优控制器,实现对伪线性系统的控制,保证飞行器纵向运动稳定和高精度控制。

首先,定义以下形式的非线性坐标变换

$$\begin{cases} \xi_1 = e_\varphi \\ \xi_2 = \dot{e}_\varphi \end{cases}$$

(15)

其中, $e_\varphi = \varphi - \varphi_c$, φ_c 为飞行器俯仰角指令。对上式进行微分得

$$\dot{\xi} = A\xi + BW$$

(16)

其中

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, W = \ddot{\varphi} - \ddot{\varphi}_c$$

(17)

为了得到 W ,可采用最优状态调节器。因为最优状态调节器的解具有统一的解析表达式,且可引入一个简单的状态线性反馈控制律,便于计算和实现反馈控制,另外,最优状态调节器除易于实现、具有工程性以外,还具有良好的频响特性,可以实现极点的最优配置^[4]。 W 的最优控制表

达式为^[4]

$$W = -r^{-1}B^TP\xi$$

(18)

其中对称非负矩阵 P 满足黎卡提矩阵代数方程

$$A^TP + PA - r^{-1}PBB^TP + Q = 0$$

(19)

式中, Q 为选定的非负定对称矩阵, r 为选定的正实数。

综上可得纵向逆控制系统由内环动态逆控制器和外环最优调节器组成,控制系统结果如图1所示。



图1 纵向逆控制系统结构框图
Fig. 1 Longitudinal inversion control system structure

4 控制系统仿真分析

纵向控制系统的输出为俯仰角,故最直接的控制方式是跟踪俯仰角指令,而在实际中,控制指令一般以攻角形式给出。俯仰角和攻角可由欧拉角关系式(2)相互转化,故两种控制指令本质上是一致的。俯仰角变化率指令用式(20)计算

$$\dot{\varphi}_c = \frac{\varphi_c - \varphi}{K_\omega \Delta t}$$

(20)

其中 Δt 为仿真时间步长; K_ω 为俯仰角变化率指令参数;俯仰角二阶导数指令为 $\ddot{\varphi}_c = 0$ 。

仿真步长取为50ms。仿真初始条件如表1所列。

表1 仿真初始条件				
Tab. 1 Initial parameters in simulation				
变量	V_0 (m/s)	H_0 (km)	x_0 (km)	y_0 (km)
数值	6000	60	160	58
变量	ω_{s0} (rad/s)	θ_0 (°)	φ_0 (°)	α_0 (°)
数值	0	0	15	15

控制器相关参数取为: $Q = \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix}$, $r = 1$, $K_\omega = 30$ 。仿真中考虑如下参数不确定性

$$\begin{cases} \rho = \rho_0(1 + \Delta\rho) \\ J_z = J_{z0}(1 + \Delta J) \\ m_z = m_{z0}(1 + \Delta m_s) \end{cases}$$

(21)

其中 $|\Delta\rho| \leq 0.3$, $|\Delta J_z| \leq 0.1$, $|\Delta m_z| \leq 0.35$ 。同时由于高空中的横向风等因素的作用,临近空间飞行器会受到外界干扰作用,仿真中认为外界干扰力矩为一随机量,且其均方差为控制力矩的20%。

仿真中还加入考虑舵机系统的实际工作特

性,主要包括:幅值限制 $\pm 20^\circ$;指令延迟 10ms;超调量 5%;最大角速度 $100^\circ/\text{s}$ (相当于 $5^\circ/\text{控制周期}$)。分别采用俯仰角指令 $\varphi_c = 10^\circ$ 和攻角指令 $\alpha_c = 10^\circ$ 进行仿真计算,仿真结果分别如图 2 和图 3 所示。

由图 2 和图 3 可知:在存在参数不确定性和外界干扰作用,同时考虑舵机实际工作特性的条件下,纵向逆控制系统可以准确跟踪俯仰角或攻角指令,俯仰角控制精度为 0.006° ,攻角控制精度为 0.036° ,可以满足控制精度要求。可见,基于动态逆和最优调节器的纵向控制系统对参数不确定性和外界干扰具有较好的鲁棒性,可以适应

舵机实际工作特性,具有工程适用性。

5 结 论

高超声速飞行器无动力滑翔阶段纵向运动模型是可逆的,系统相对阶等于绕质心运动模型的系统阶数,应用动态逆方法可实现绕质心运动模型的线性化。

由内环动态逆控制器和外环最优调节器所组成的飞行器纵向逆控制系统,在考虑舵机实际工作特性的条件下可以准确跟踪俯仰角或攻角指令,满足控制精度要求,对参数不确定性和外界干扰具有较好的鲁棒性,具有工程适用性。

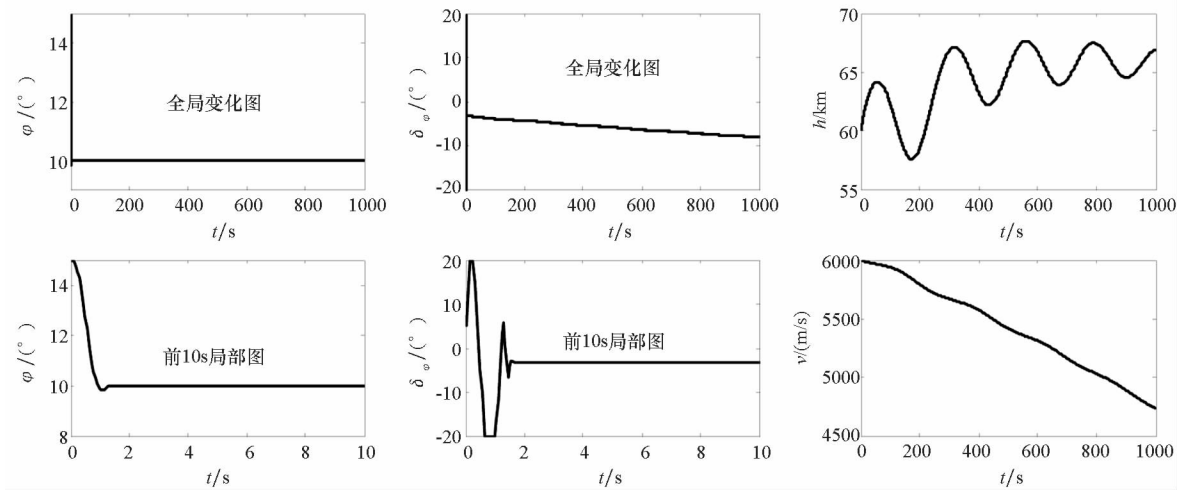


图 2 俯仰角指令控制结果及其局部放大图
Fig. 2 Results under the pitch angle command and local enlargements

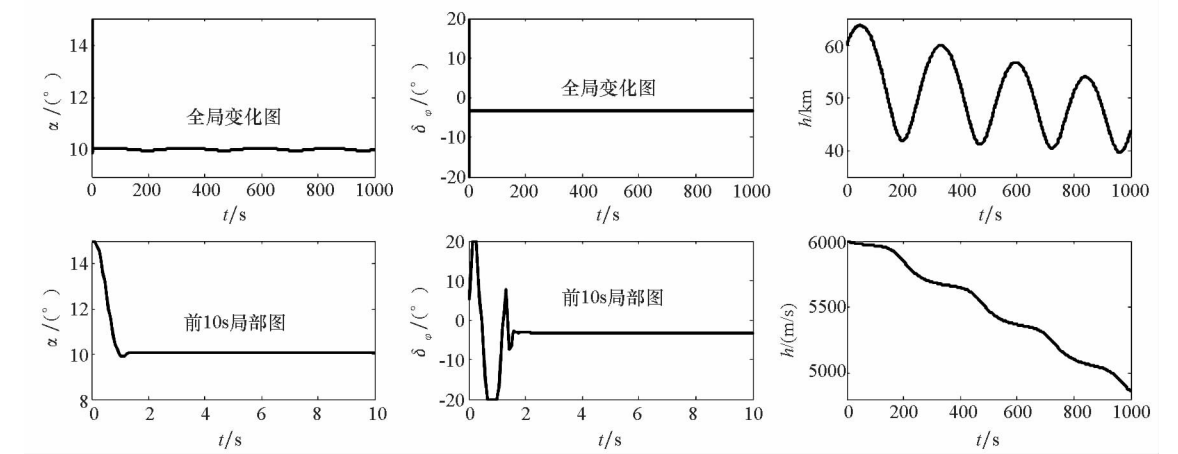


图 3 攻角指令控制结果及其局部放大图
Fig. 3 Results under the angle of attack command and local enlargements

and characterization [J]. Journal of Applied physics, 1998 , 84(7) : 3912 - 3916.

[4] Xue S W, Zu X T, Zheng X G, et al. Effect of annealing and dopant concentration on the optical characteristics of ZnO: Al thin films by sol-gel technique [J]. physic B: Condensed Matter, 2006, 382: 201 - 204.

[5] Valle G G, Hammer P, et al. Transparent and conductive ZnO: Al thin films prepared by Sol-gel dipcoating [J]. Journal of the European Ceramic Society, 2004, 24: 1009 - 1013.

[6] 尹玉刚, 沈鸿烈, 楼晓波, 等. 溶胶 - 凝胶法生长(002)高度择优取向的 ZnO: Al 薄膜[J]. 功能材料, 2008(7): 1122 - 1125.

YIN Yugang, SHEN Honglie, LOU Xiaobo, et al. Highly (002) oriented ZnO: Al films prepared by Sol-gel method [J]. Journal of Functional Materials, 2008(7) : 1122 - 1125. (in Chinese)

[7] 倪星元, 姚兰芳, 沈军, 等. 纳米材料制备技术[M]. 北京: 化学工业出版社, 2008.

NI Xingyuan, YAO Lanfang, SHEN Jun, et al. The preparation technology of Nano-materials [M]. Beijing: Chemical Industry Press, 2008. (in Chinese)

[8] Zhao L, Lian J S, Liu Y H, et al. Structural and optical properties of nanostructured ZnO thin films deposited on quartz glass [J]. Applied Surface Science, 2006, 252: 8451 - 8455.

[9] Keh M L, Pajjay T. Growth mechanism and characterization of ZnO: Al multi-layered thin films by sol-gel technique [J]. Thin Solid Films, 2007, 515: 8601 - 8604.

[10] Lu W L, Huang K C, Hung P K, et al. Study of textured ZnO: Al thin film and its optical properties for thin film silicon solar cells [J]. Journal of Physics and Chemistry of Solids, 2012, 73(1) : 52 - 56

[11] Lou X B, Shen H L, Zhang H, et al. Optical properties of nanosized ZnO films prepared by sol - gel process [J]. Transactions Nonferrous met Soc China, 2007, 17: 814 - 817.

[12] 高立, 张建民. 带隙可调的 Al, Mg 掺杂 ZnO 薄膜的制备[J]. 物理学报, 2009, 58(10) : 7199 - 7203.

GAO Li, ZHANG Jianmin. Preparation of Mg and Al co-doped ZnO thin films with tunable band gap [J]. Acta Physica Sinica, 2009, 58(10) : 7199 - 7203. (in Chinese)

[13] 张天宝, 李金培. 制备条件对铝掺杂氧化锌(AZO)薄膜结构和电学性质的影响[J]. 影像科学与光化学, 2009, 27(6) : 452 - 461.

ZHANG Tianbao, LI Jinpei. Effects of preparing conditions on structure and electrical properties of AZO films [J]. Imaging Science and Photochemistry, 2009, 27(6) : 452 - 461. (in Chinese)

[14] Jiang X, Wong F L, Fung M K, et al. Aluminum-doped zinc oxide films as transparent conductive electrode for organic light-emitting devices [J]. Applied Physics Letters, 2003, 83: 1875 - 1879.

[15] Dieter K S. 半导体材料与器件表征技术[M]. 刘爱民, 张贺秋, 刘艳红, 等译. 大连: 大连理工大学出版社, 2008.

Dieter K S. Semiconductor material and device characterization [M]. LIU Aimin, ZHANG Heqiu, LIU Yanhong, et al. Dalian: Dalian University of Technology Press, 2008. (in Chinese)

(上接第 32 页)

参考文献 (References)

[1] Snell S A, Enns D F, Garrard L. Nonlinear inversion flight control for a super maneuverable aircraft [J]. Journal of Guidance, Control, and Dynamics, 1992, 15(4) : 976 - 984.

[2] Wang Q, Stengel R F. Robust nonlinear control of a hypersonic aircraft [J]. Journal of Guidance, Control, and Dynamics, 2000, 23(4) : 577 - 585.

[3] 张友安, 胡云安, 苏身榜. BTT 导弹控制系统鲁棒动态逆设计[J]. 宇航学报, 2002, 32(2) : 89 - 91.

ZHANG Youan, HU Yunan, SU Shenbang. Robust dynamic inversion control for BTT missile[J]. Journal of Astronautics, 2002, 32(2) : 89 - 91. (in Chinese)

[4] 刘燕斌, 陆宇平. 非线性动态逆控制在高超飞控系统中的应用[J]. 应用科学学报, 2006, 24(6) : 613 - 617.

LIU Yanbin, LU Yuping. Application of nonlinear dynamic inversion control in flight control of hypersonic vehicle [J]. Journal of Applied Sciences, 2006, 24(6) : 613 - 617. (in Chinese)

[5] Lee H P, Reiman S E, Dillon C H. Robust nonlinear dynamic inversion control for a hypersonic cruise vehicle [R]. AIAA 2007 - 6685, 2007.

[6] Kawaguchi J, Miyazawa Y. Stochastic evaluation and optimization of the hierarchy-structured dynamic inversion flight control [R]. AIAA 2009 - 6175, 2009.

[7] 武立军, 贺有智. 应用时标分离和动态逆方法设计飞行器的姿态控制系统[J]. 现代防御技术, 2007, 35(4) : 55 - 58.

WU Lijun, HE Youzhi. Design method to attitude control system of flying vehicle with dynamic inversion and time scale separation [J]. Modern Defence Technology, 2007, 35(4) : 55 - 58. (in Chinese)

[8] 戴先中. 多变量非线性系统的神经网络逆控制方法[M]. 北京: 科学出版社, 2005: 29 - 45.

DAI Xianzhong. Neural network inverse control method for multivariable nonlinear system [M]. Beijing: Science Press, 2005: 29 - 45. (in Chinese)

具有 C 轴取向 Al^{3+} 掺杂型 ZnO 薄膜的 溶胶 - 凝胶法制备及其性能研究*

郑春满, 宋植彦, 魏海博, 帖楠, 谢凯
(国防科技大学 航天与材料工程学院, 湖南 长沙 410073)

摘要: 以乙二醇甲醚为溶剂, 采用 Sol-Gel 法制备出具有 C 轴取向、可导电的 Al^{3+} 离子掺杂 ZnO 透明薄膜, 并利用场发射扫描电镜、X - 射线衍射、能谱分析、标准四探针和反射光谱仪等对薄膜的组成、结构和光学性能进行了分析。结果表明: Al^{3+} 离子掺杂 ZnO 薄膜为六方纤锌矿型结构, 由六棱柱状阵列构成, 具有 C 轴择优取向; 薄膜电阻率随 Al^{3+} 离子掺杂浓度的升高而降低; 在可见光区域, 薄膜透光率随 Al^{3+} 离子掺杂浓度的升高而降低, 掺杂 3% ZnO 薄膜的透光率达到 90% 左右, 禁带宽度为 3.25 eV, 具备制作薄膜太阳能电池透明导电电极材料的应用价值。

关键词: 溶胶 - 凝胶; ZnO 薄膜; Al^{3+} 离子掺杂; 制备

中图分类号: O648 文献标志码: A 文章编号: 1001 - 2486(2012)03 - 0033 - 05

The study on the preparation of Al^{3+} doped ZnO thin films with [002] oriented by sol-gel method and its properties

ZHENG Chunman, SONG Zhiyan, WEI Haibo, TIE Nan, XIE Kai

(College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The aluminum-doped ZnO thin films with [002] oriented were prepared using 2-methoxyethan as a solvent system by the sol-gel method. The composition, structure and optical properties were studied by means of Scanning Electron Microscopy (SEM), X-ray Diffraction (XRD), Energy-dispersive X-ray microanalysis (EDX) and Spectral shape. The results showed that the ZnO films are hexagonal wurtzite structure, which consists of hexagonal rods growing along C axis. The resistivity of the aluminum-doped ZnO thin films decreases with the increase of the content of Al^{3+} . In the visible region, the light transmittance of the ZnO thin films with 3% aluminum is about 90%, and the band gap is about 3.25 eV, which can meet the material requirements for photoelectric devices such as photovoltaic solar cell.

Key words: Sol-Gel; ZnO thin film; Aluminum-doped; preparation

氧化锌(ZnO)薄膜作为 II - VI 族宽带隙半导体材料, 具有低介电常数、高化学稳定性、优良的光电和压电特性以及机电耦合性等。同时, ZnO 薄膜及器件的制备工艺可与硅微电子集成工艺相容, 易于实现光电集成。因此, ZnO 薄膜在众多领域有着广泛的应用, 如用作太阳能电池的透明导电膜和过渡层, 用于压电器件、平板显示器的透明电极、电磁防护屏建筑玻璃的红外反射涂层以及传感器等方面^[1-3]。

与纯相 ZnO 薄膜相比, Al^{3+} 离子掺杂 ZnO 薄膜具有一系列优点而成为国内外研究热点^[4-6]。(1) 宽禁带值一般大于 3.1 eV (可见光光子的能量), 具有紫外截止特性; (2) 可见光区透射率高、对红外光区具有高反射率。可见光区平均透射率约为 85%, 红外光区反射率约为 80%; (3) 导电性能好, 直流电阻率最低可达 $10^{-4} \Omega \cdot \text{cm}$ 数量

级; (4) 载流子浓度 n 值较大, 可达 $10^{19} \sim 10^{21} \text{ cm}^{-3}$; (5) 霍尔迁移率通常在 $10 \sim 40 \text{ cm}^2 \text{ V}^{-1} \cdot \text{S}^{-1}$ 。

目前, Al^{3+} 离子掺杂的 ZnO 薄膜的制备方法很多, 主要有磁控溅射法^[7]、喷雾热分解法^[7]、金属有机物化学气相沉积法^[7-8]、脉冲激光沉积法^[7]、激光分子束外延法^[7]、溶胶 - 凝胶法^[7, 9]、化学气相沉积法^[7]等。与其他方法相比较, 溶胶 - 凝胶法^[7, 9]具有成膜均匀性好, 与衬底附着力强, 易于原子级掺杂, 可精确控制掺杂水平等优点, 而且无需真空设备, 工艺简单, 可获得理想厚度和组分的薄膜。如 Xue 等^[10]采用溶胶凝胶法制备了 Al^{3+} 离子掺杂的 ZnO 薄膜, 电阻率最低可达 $3.7 \times 10^{-4} \Omega \cdot \text{cm}$, 在可见光区的透过率达到 90% 以上; Lou 等^[11]以二乙醇胺为溶剂, 采用溶胶凝胶法制备了 ZnO 薄膜, 其禁带宽度为 3.265

* 收稿日期: 2011 - 06 - 22

基金项目: 湖南省自然科学基金资助项目(10JJ4045); 国防科技大学校预研资助项目(JC08 - 01 - 06)

作者简介: 郑春满(1976—), 男, 山东平度人, 副教授, 博士, E-mail: zhengchunman@hotmail.com

~3.293 eV。

本文主要以乙二醇甲醚为溶剂,采用溶胶-凝胶方法制备 Al^{3+} 掺杂型的 ZnO 薄膜,利用一系列的分析测试手段系统地对薄膜的组成、结构和光学性能进行了分析。

1 实验

1.1 Al^{3+} 掺杂 ZnO 薄膜的制备

利用预先制备的胶体,通过 Dip-coating 方法在已清洗洁净的载玻片上制备 Al^{3+} 掺杂型 ZnO 薄膜。溶胶由二水合醋酸锌 ($\text{Zn}(\text{Ac})_2 \cdot 2\text{H}_2\text{O}$, 99.0%)、乙二醇甲醚(溶剂)、单乙醇胺(稳定剂)以及不同浓度的六水合氯化铝 ($\text{AlCl}_3 \cdot 6\text{H}_2\text{O}$, 99.0%) 制得。其中, Zn^{2+} 物质的量浓度为 $0.5 \text{ mol} \cdot \text{L}^{-1}$, 单乙醇胺和 Zn^{2+} 物质的量之比为 1:1, 铝离子的浓度与锌离子摩尔比分别为 1.0%、3.0% 和 5.0%。混合溶液在 60°C 下搅拌 8 h 变澄清, Dip-coating 制备的薄膜置于 90°C 烘箱中烘干,通过重复 Dip-coating 操作制备具有不同厚度的薄膜。膜的预处理温度为 310°C , 最终高温处理温度为 600°C , 恒温时间为 10 min, 升温速率为 $100^\circ\text{C}/\text{h}$ 。

1.2 测试与表征

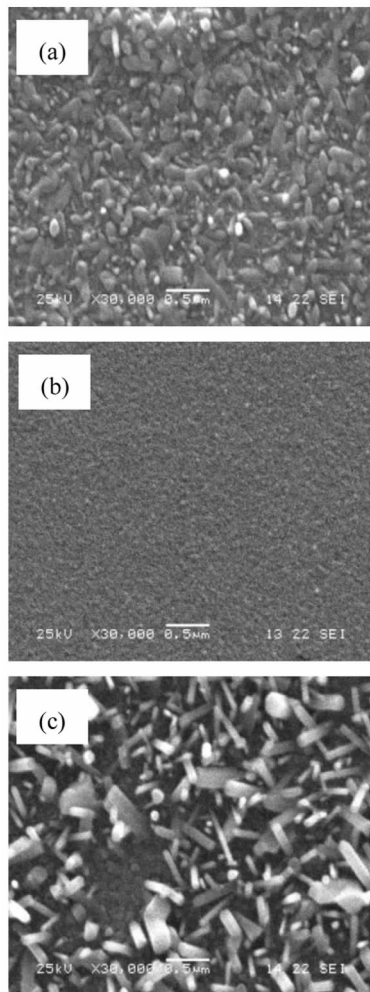
采用日本 HITACHI S4800 场发射扫描电镜观察样品的表面形貌; 采用德国布鲁克 D8 ADVANCED 型 X-射线衍射仪, 以 $\text{CuK}\alpha$ 为 X 射线源, 扫描步长为 $0.02^\circ/0.15\text{s}$, 对所得产物的物相与其结构进行测定; 采用 SX1934 (SZ282) 型数字式四探针测试仪测量薄膜的方块电阻并计算薄膜的电阻率; 采用美国 PkintElmer 公司的 Lambda 900 反射光谱仪对薄膜进行光谱测量, 采样间隔为 1 nm, 光谱范围 300 ~ 2000 nm。

2 结果与讨论

2.1 Al^{3+} 掺杂 ZnO 薄膜的形貌、组成与结构分析

在 ZnO 薄膜中, 组成、结构和形貌都是影响其电学和光学性能的重要因素。图 1 是不同 Al^{3+} 掺杂浓度下所制备的 ZnO 薄膜的 SEM 图。可以看出, Al^{3+} 掺杂浓度为 1.0% 的 ZnO 颗粒堆积紧密, 形成了均匀致密、表面呈现为不完整的六角柱状分布的薄膜; 掺杂浓度为 3.0% 的氧化锌薄膜在 600°C 热处理后, 薄膜中氧化锌的晶粒非常细小, 说明随着铝掺杂浓度的升高, 高温下氧化锌晶粒的生长受到了明显抑制; 但是, 随着 Al^{3+}

掺杂浓度进一步提高, 如图 1(c) 所示, ZnO 薄膜结晶度进一步提高, 出现了垂直于衬底的具有较好结晶性能 ZnO 纳米晶, 对于这一现象形成原因将在下面进行详细分析。



(a) Al^{3+} 掺杂浓度 1.0%

(b) Al^{3+} 掺杂浓度 3.0%

(c) Al^{3+} 掺杂浓度 5.0%

图 1 不同 Al^{3+} 掺杂浓度所制备 ZnO 薄膜 SEM 图

Fig. 1 SEM of ZnO films with different Al^{3+} concentration

在溶胶制备过程中, 铝元素是否掺杂进入 ZnO 晶格是掺杂的关键。图 2 是不同 Al^{3+} 掺杂浓度下所制备的 ZnO 薄膜的 XRD 谱图。从图中可以看出, 2θ 为 31.49° 、 34.41° 、 35.86° 、 47.54° 处出现了分别归属于 (100)、(002)、(101) 和 (102) 晶面的氧化锌的衍射峰^[12-13], 但并未出现归属于氧化铝的衍射峰。这说明: (1) Al^{3+} 掺杂 ZnO 薄膜具有与 ZnO 相同的晶体结构, 即六方纤锌矿结构。而 $2\theta = 34.41^\circ$ 的衍射峰为 (002) 晶面衍射峰, 表明样品具有 (002) 晶向择优取向, 即纤锌矿结构 C 轴择优取向生长。这与图 1 中的 SEM 分析相一致; (2) 虽然 Al 元素进入 ZnO 薄膜中, 但

Al 元素的掺杂没有形成新的化合物 (Al₂O₃), Al³⁺ 只起掺杂替代 Zn²⁺ 作用。

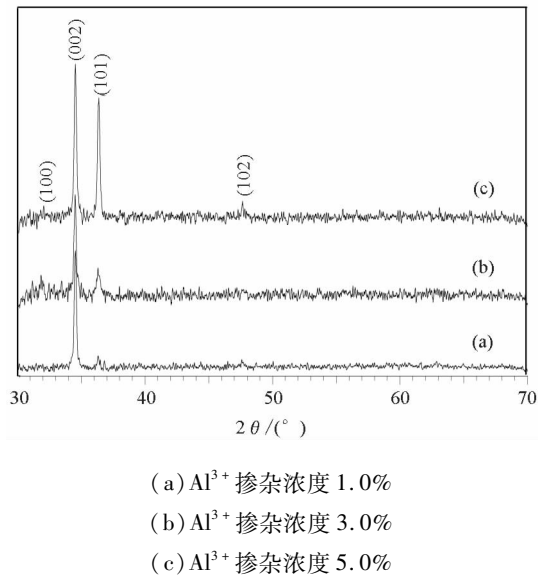


图2 不同 Al³⁺ 掺杂浓度制备 ZnO 薄膜 XRD 谱图

Fig. 2 The XRD patterns of ZnO films with different Al³⁺ concentration

同时,由图可知,与标准 ZnO 薄膜相比,由于 Al 的掺杂效应造成了晶格畸变、衍射角偏离的现象。随着铝离子掺杂浓度提高,归属于 (101) 晶面衍射峰强度逐渐增强,而且逐渐向右发生偏移。由于铝离子半径小于锌离子半径,当 Al³⁺ 在 ZnO 晶格中掺杂替代 Zn²⁺ 后,在结晶过程中会产生残余应力,即 Al³⁺ 对 Zn²⁺ 的掺杂替代必然会造成晶格畸变。根据文献[14],随着 Al³⁺ 的掺杂越多,晶格畸变现象也越严重,衍射角偏离也越大,但不会无限制偏离,因为 Al³⁺ 加入量达到一定程度时,将形成 Al₂O₃ 晶体,不起掺杂替代作用了。这也是 Al³⁺ 掺杂浓度为 5.0% 的 ZnO 薄膜出现了垂直于衬底的具有较好结晶性能的 ZnO 纳米晶的主要原因之一。

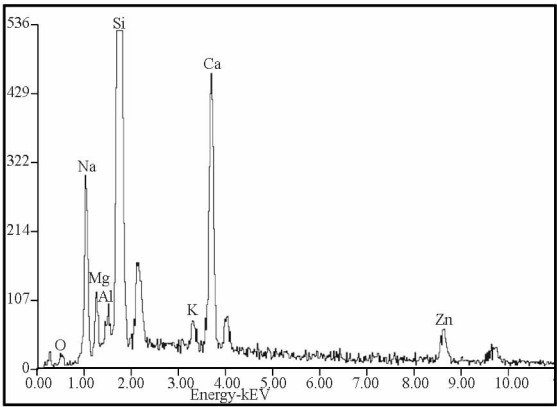


图3 Al³⁺ 掺杂浓度 3% 制备 ZnO 薄膜 EDX 分析

Fig. 3 The EDX analyse of ZnO films with 3% Al³⁺

为进一步研究 Al³⁺ 掺杂 ZnO 薄膜组成,采用 EDX 对 Al³⁺ 掺杂浓度 3.0% 制备 ZnO 薄膜进行分析,如图 3 所示。从图中可以看出,薄膜中包括锌、铝、钠、镁、硅、钙、钾、氧等多种元素,其中钠、镁、硅、钙、钾等为玻璃基底所含元素。薄膜中铝元素与锌元素的比例并非最初原料所掺杂的 3.0%。这可能是玻璃基底中含有元素铝缘故。

2.2 Al³⁺ 掺杂 ZnO 薄膜电学与光学性能分析

表 1 是不同掺铝量与 ZnO 薄膜电阻的关系。从表 1 中可以看出,纯 ZnO 薄膜电阻率最高,为 3730Ω · cm;随着 Al³⁺ 掺杂浓度的提高,薄膜电阻率下降,Al³⁺ 掺杂浓度 5.0% 时薄膜电阻率为 1.56Ω · cm。

表 1 不同掺铝量与 ZnO 薄膜电阻的关系

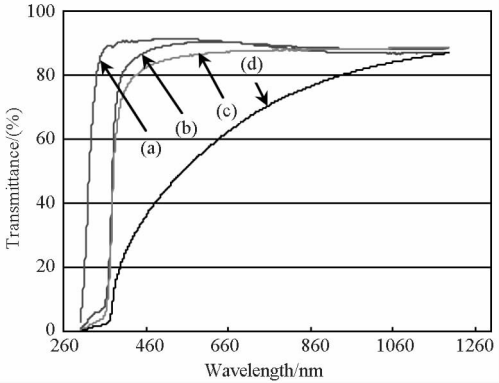
Tab. 1 The relationship of resistivity and the content of Al³⁺

样品	浓度 (mol · L ⁻¹)	掺铝量 (%)	电阻率 (Ω · cm)
A	0.5	0	3730
B	0.5	1.0	2.36
C	0.5	3.0	1.56
D	0.5	5.0	1.34

文献[6]表明,在通常条件下,ZnO 薄膜导电机制主要是通过晶界热离子的热场发射,在较高温度下,主要是电离杂质散射,而且晶体不完整性导致的严重散射使纯 ZnO 薄膜的电阻率最高。当 ZnO 薄膜中掺入 Al³⁺ 后,Al 原子趋向于以 Al³⁺ + 3e 的方式发生固溶,Al³⁺ 离子占据晶格中 Zn²⁺ 离子位置,形成一个 +1 价正电荷中心 AlZn 和一个多电子,这个多余的价电子挣脱束缚而成为导电电子,增加了薄膜中自由电子和载流子浓度,从而使薄膜电阻率降低。而且,铝离子浓度越高,薄膜中导电电子越多,其电阻率也越低。

图 4 是不同 Al³⁺ 掺杂浓度下所制备的 ZnO 薄膜的透过率光谱图。由图可见,(1)所有样品在近紫外光波段透射率急剧减小,对紫外光截止,这主要是 ZnO 薄膜的光学禁带宽度所致,薄膜的光谱吸收限约为 365 nm。(2)在可见光范围内,当 Al³⁺ 掺杂浓度小于 3.0% 时,薄膜透射率较高,平均达到 90% 左右,并且吸收边陡峭,表明薄膜结晶质量良好,均匀致密,呈现典型宽禁带半导体薄膜的光学特征,这种具有高透光率 ZnO 薄膜在用作窗口材料方面具有很大潜力;当 Al³⁺ 掺杂浓度为 5.0% 时,薄膜在可见光区的透射率较其他样品有较大降低,结合前述 SEM 和 XRD 分析,这

与样品表面形成垂直于衬底的具有较好结晶性能 ZnO 纳米晶有关,此时薄膜中缺陷、晶界等成分不均匀处也大大增加,造成光的散射,使得散射因子 S 和吸收系数 α 同时增加,导致薄膜透射率下降。

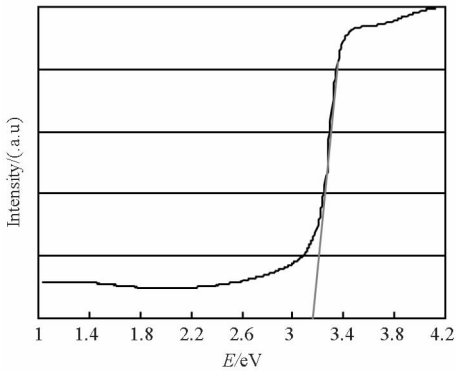


(a) 空白基底
(b) Al^{3+} 掺杂浓度为 1.0%
(c) Al^{3+} 掺杂浓度为 3.0%
(d) Al^{3+} 掺杂浓度为 5.0%

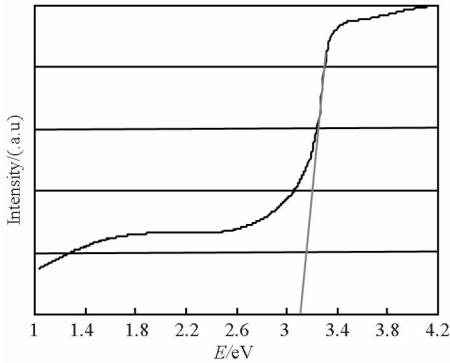
图 4 不同 Al^{3+} 掺杂浓度下所制备的 ZnO 薄膜的透过率光谱图
Fig. 4 The spectrogram of ZnO films with different Al^{3+} concentration

由光谱吸收理论可知,当光子能量大于禁带宽度 E_g 以后,吸收曲线将急剧上升,表示有强烈光吸收,这一吸收系数快速上升过程对应电子的直接跃迁过程。根据透过曲线可计算出薄膜吸收曲线,从而推算 ZnO 薄膜的禁带宽度,并计算出本征吸收限^[15]。

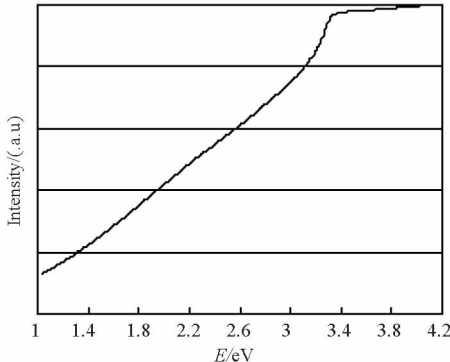
图 5 是不同 Al^{3+} 掺杂浓度下所制备的 ZnO 薄膜的吸收谱图。从图中可以看出,掺杂浓度为 1.0% 和 3.0% 的 ZnO 薄膜的禁带宽度均在 3.2 eV 左右,这与其它报道的 ZnO 材料的禁带宽度 3.30 eV 基本符合。当 Al^{3+} 掺杂浓度为 5.0% 时,薄膜的吸收曲线并未出现急剧上升,这与该样品在制备过程中所形成的组成和结构有关。



(a) Al^{3+} 掺杂浓度为 1.0%



(b) Al^{3+} 掺杂浓度为 3.0%



(c) Al^{3+} 掺杂浓度为 5.0%

图 5 不同 Al^{3+} 掺杂浓度下所制备的 ZnO 薄膜的吸收谱图
Fig. 5 The absorbance of ZnO films with different Al^{3+} concentration

3 结 论

以乙二醇甲醚为溶剂,单乙醇胺为稳定剂,采用 Sol-Gel 法制备出了具有高度 C 轴取向性、可导电的 Al^{3+} 离子掺杂的 ZnO 透明薄膜。 Al^{3+} 离子掺杂浓度为 3% 时所制备的 ZnO 薄膜的电阻率均值为 $1.56 \Omega \cdot \text{cm}$,其结构为六方纤锌矿型,由六棱柱状阵列构成,具有高度的 C 轴择优取向;在可见光区域透光率达到 90% 左右,禁带宽度为 3.2 eV 左右,具备制作薄膜太阳能电池透明导电电极材料的应用价值。

参考文献 (References)

[1] Atsushi T, Akira O, Takeyoshi O, et al. Repeated temperature modulation epitaxy for p-type doping and light-emitting diode based on ZnO [J]. Nature Materials, 2005, 4: 42-46.
[2] 邢光建, 李钰梅, 江伟, 等. ZnO 薄膜的 p 型掺杂研究进展[J]. 真空, 2009, 49(4): 41-48.
XING Guangjian, LI Yumei, JIANG Wei, et al. Recent progress in R&D of p-type doping of ZnO thin films [J]. Vacuum, 2009, 49(4): 41-48. (in Chinese)
[3] Chen Y, Bagnall D M, Koh H J, et al. Plasma assisted molecular beam epitaxy of ZnO on c-plane sapphire: growth

and characterization [J]. Journal of Applied physics, 1998 , 84(7) : 3912 - 3916.

[4] Xue S W, Zu X T, Zheng X G, et al. Effect of annealing and dopant concentration on the optical characteristics of ZnO: Al thin films by sol-gel technique [J]. physic B: Condensed Matter, 2006, 382: 201 - 204.

[5] Valle G G, Hammer P, et al. Transparent and conductive ZnO: Al thin films prepared by Sol-gel dipcoating [J]. Journal of the European Ceramic Society, 2004, 24: 1009 - 1013.

[6] 尹玉刚, 沈鸿烈, 楼晓波, 等. 溶胶 - 凝胶法生长(002)高度择优取向的 ZnO: Al 薄膜[J]. 功能材料, 2008(7): 1122 - 1125.

YIN Yugang, SHEN Honglie, LOU Xiaobo, et al. Highly (002) oriented ZnO: Al films prepared by Sol-gel method [J]. Journal of Functional Materials, 2008(7) : 1122 - 1125. (in Chinese)

[7] 倪星元, 姚兰芳, 沈军, 等. 纳米材料制备技术[M]. 北京: 化学工业出版社, 2008.

NI Xingyuan, YAO Lanfang, SHEN Jun, et al. The preparation technology of Nano-materials [M]. Beijing: Chemical Industry Press, 2008. (in Chinese)

[8] Zhao L, Lian J S, Liu Y H, et al. Structural and optical properties of nanostructured ZnO thin films deposited on quartz glass [J]. Applied Surface Science, 2006, 252: 8451 - 8455.

[9] Keh M L, Pajjay T. Growth mechanism and characterization of ZnO: Al multi-layered thin films by sol-gel technique [J]. Thin Solid Films, 2007, 515: 8601 - 8604.

[10] Lu W L, Huang K C, Hung P K, et al. Study of textured ZnO: Al thin film and its optical properties for thin film silicon solar cells [J]. Journal of Physics and Chemistry of Solids, 2012, 73(1) : 52 - 56

[11] Lou X B, Shen H L, Zhang H, et al. Optical properties of nanosized ZnO films prepared by sol - gel process [J]. Transactions Nonferrous met Soc China, 2007, 17: 814 - 817.

[12] 高立, 张建民. 带隙可调的 Al, Mg 掺杂 ZnO 薄膜的制备[J]. 物理学报, 2009, 58(10) : 7199 - 7203.

GAO Li, ZHANG Jianmin. Preparation of Mg and Al co-doped ZnO thin films with tunable band gap [J]. Acta Physica Sinica, 2009, 58(10) : 7199 - 7203. (in Chinese)

[13] 张天宝, 李金培. 制备条件对铝掺杂氧化锌(AZO)薄膜结构和电学性质的影响[J]. 影像科学与光化学, 2009, 27(6) : 452 - 461.

ZHANG Tianbao, LI Jinpei. Effects of preparing conditions on structure and electrical properties of AZO films [J]. Imaging Science and Photochemistry, 2009, 27(6) : 452 - 461. (in Chinese)

[14] Jiang X, Wong F L, Fung M K, et al. Aluminum-doped zinc oxide films as transparent conductive electrode for organic light-emitting devices [J]. Applied Physics Letters, 2003, 83: 1875 - 1879.

[15] Dieter K S. 半导体材料与器件表征技术[M]. 刘爱民, 张贺秋, 刘艳红, 等译. 大连: 大连理工大学出版社, 2008.

Dieter K S. Semiconductor material and device characterization [M]. LIU Aimin, ZHANG Heqiu, LIU Yanhong, et al. Dalian: Dalian University of Technology Press, 2008. (in Chinese)

(上接第 32 页)

参考文献 (References)

[1] Snell S A, Enns D F, Garrard L. Nonlinear inversion flight control for a super maneuverable aircraft [J]. Journal of Guidance, Control, and Dynamics, 1992, 15(4) : 976 - 984.

[2] Wang Q, Stengel R F. Robust nonlinear control of a hypersonic aircraft [J]. Journal of Guidance, Control, and Dynamics, 2000, 23(4) : 577 - 585.

[3] 张友安, 胡云安, 苏身榜. BTT 导弹控制系统鲁棒动态逆设计[J]. 宇航学报, 2002, 32(2) : 89 - 91.

ZHANG Youan, HU Yunan, SU Shenbang. Robust dynamic inversion control for BTT missile[J]. Journal of Astronautics, 2002, 32(2) : 89 - 91. (in Chinese)

[4] 刘燕斌, 陆宇平. 非线性动态逆控制在高超飞控系统中的应用[J]. 应用科学学报, 2006, 24(6) : 613 - 617.

LIU Yanbin, LU Yuping. Application of nonlinear dynamic inversion control in flight control of hypersonic vehicle [J]. Journal of Applied Sciences, 2006, 24(6) : 613 - 617. (in Chinese)

[5] Lee H P, Reiman S E, Dillon C H. Robust nonlinear dynamic inversion control for a hypersonic cruise vehicle [R]. AIAA 2007 - 6685, 2007.

[6] Kawaguchi J, Miyazawa Y. Stochastic evaluation and optimization of the hierarchy-structured dynamic inversion flight control [R]. AIAA 2009 - 6175, 2009.

[7] 武立军, 贺有智. 应用时标分离和动态逆方法设计飞行器的姿态控制系统[J]. 现代防御技术, 2007, 35(4) : 55 - 58.

WU Lijun, HE Youzhi. Design method to attitude control system of flying vehicle with dynamic inversion and time scale separation [J]. Modern Defence Technology, 2007, 35(4) : 55 - 58. (in Chinese)

[8] 戴先中. 多变量非线性系统的神经网络逆控制方法[M]. 北京: 科学出版社, 2005: 29 - 45.

DAI Xianzhong. Neural network inverse control method for multivariable nonlinear system [M]. Beijing: Science Press, 2005: 29 - 45. (in Chinese)

双酚 A 环氧树脂改性氧化双环戊二烯体系
固化动力学分析及其性能研究*

吕文丽^{1,2}, 肖加余¹, 徐旺生², 曾竟成¹, 刘 钧¹, 蒋卫和³

(1. 国防科技大学 航天与材料工程学院, 湖南 长沙 410073;

2. 武汉工程大学 化工与制药学院, 湖北 武汉 430073;

3. 湖南岳阳昌德化工实业有限公司, 湖南 岳阳 414004)

摘 要: 氧化双环戊二烯(CDR-0122)环氧树脂固化物的耐热性较好,但脆性大。采用不同质量分数的环氧(E-51)改性 CDR-0122。对不同的改性体系进行 DSC 分析,确定其活化能、反应级数和最佳固化工艺。对改性的 CDR-0122 环氧树脂固化物进行力学性能和耐温性能测试。结果表明:当加入 E-51 的质量分数为 30% 时,CDR-0122 的树脂体系的综合性能最优。相比改性前,改性后的树脂体系活化能略有下降,并且其拉伸强度和弯曲强度分别提高了 52.38% 和 54.77%,但玻璃化转变温度略有下降。

关键词: 氧化双环戊二烯; E-51; 改性; 力学性能; 耐温性能;

中图分类号: TN248 文献标志码: A 文章编号: 1001-2486(2012)03-0038-06

Curing dynamics analysis and properties of acycloaliphatic modified
by bisphenol A(E-51) epoxy resin

LV Wenli^{1,2}, XIAO Jiayu¹, XU Wangsheng², ZENG Jingcheng¹, LIU Jun¹, JIANG Weihe³

(1. College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China;

2. College of Chemical Engineering and Pharmacy, University of Wuhan Institute of Technology, Wuhan 430073, China;

3. Changde Chemical Corporation Limited, Yueyang 414004, China)

Abstract: Acycloaliphatic(CDR-0122) epoxy resin has good heat-resistance, but it is characterized by being brittle. CDR-0122 modified with different mass fraction of epoxy (E-51) were studied. First, DSC analysis was carried out to determine the activation energy, reaction order and the best curing procedure of the modified resin. Then mechanical property and heat-resistance property of the cured modified resin were tested. The results showed that, compared with the unmodified CDR-0122, the activation energy of the modified resin system decreased slightly, and its tensile strength and flexural strength increased by 52.38% and 54.77%, respectively, but the glass transition temperature decreased slightly. Considering all aspects of performance, the system of CDR-0122 modified with 30 wt% E-51 is the best one.

Key words: acycloaliphatic; epoxy; modification; mechanics property; heat-resistance

随着近代科学技术的发展,运载火箭、导弹和航天飞机等空间运载工具以及飞机、汽车和船舶等交通工具都朝着可靠性好、寿命长和能耗低的方向发展。这些新的设计思想对树脂基复合材料的性能,特别是耐高温性能提出了更高的要求^[1]。作为复合材料基体的树脂除了要具备耐高温性能外,还需要良好的力学性能、耐疲劳和耐冲击性能等。单纯的耐热性好而其他综合性能不好的树脂,其使用价值不大。

氧化双环戊二烯(CDR-0122)是一种特种耐高温树脂,但同时 CDR-0122 环氧树脂的固化物脆性较大,限制了它的广泛应用。因此,协调

好环氧树脂固化体系韧性和耐高温性之间的矛盾,在保证它耐热性变化不大或提高的情况下增强其韧性,对扩大其应用领域具有很大的实用意义。现阶段,国内对 CDR-0122 环氧树脂的应用研究比较少,而国外已广泛应用于军工、航空等关键领域,但核心技术保密,公开的文献较少。

环氧树脂基体的增韧方法主要有两种^[2]:①在环氧树脂中加入橡胶弹性体分散相来增韧。②用热固性树脂连续贯穿于环氧树脂网络中形成互穿网络、半互穿网络结构来增韧。其中橡胶弹性体增韧改性环氧树脂是以牺牲耐热性和刚性为代价。本文采用第二种方法,通过热固性环氧树脂

* 收稿日期:2011-07-16

基金项目:国家部委基金项目

作者简介:吕文丽(1986—),女,湖北荆州人,博士研究生, E-mail: 276191477@qq.com;

肖加余(通信作者),男,教授,博士,博士生导师, E-mail: jiayuxiao@tom.com

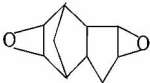
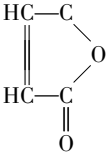
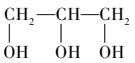
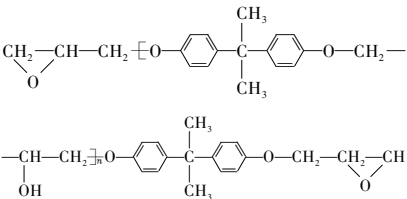
E-51 来改性 CDR-0122,由于两者同为环氧树脂,它们的相容性很好,且 E-51 引入的羟基增强其交联反应提供强的韧性并维持耐热性基本不变。同时树脂的交联程度和固化效果直接决定着固化物的性能,因此,在树脂和固化剂体系确定的前提下,选取最佳固化工艺条件以获得具有最佳固化性能的研究就显得尤为重要。树脂的固化反应动力学研究是确定其最佳固化工艺条件较为直接和有效的途径。为此,本文利用不同配比的 E-51 改性 CDR-0122 环氧树脂,并对树脂体系的固化反应动力学进行研究,以此来确定其最佳的固化工艺条件。然后测试浇铸体的拉伸性能、弯曲性能以及玻璃化转变温度。

1 实验部分

1.1 原材料

CDR-0122 环氧值为 1.20~1.22,白色粉末,产自湖南岳阳昌德化工有限公司;E-51 环氧值为 0.51~0.54,淡黄色粘稠液体,产自湖南岳阳石油化工总厂;顺酐纯度≥99.5%,分析纯,白色粉末,产自天津博迪化工股份有限公司;甘油纯度≥99.0%,分析纯,无色透明粘稠液体,产自天津市大茂化学试剂厂。结构式见表 1。

表 1 原材料结构式
Tab. 1 Chemical structures of materials

Name	Chemical structures
CDR-0122	
顺酐	
甘油	
E-51	

1.2 样品制备和测试方法

1.2.1 浇铸体制备

将环氧树脂 CDR-0122 与固化剂顺酐、促进剂甘油按质量比 100:51:7.5 混合后,在 60℃ 下

加热溶解、减压脱泡。然后加入 E-51 与固化剂顺酐的混合物,其中 E-51 环氧树脂改性剂分别按质量分数为 10%、20%、30%、40% 和 50% 加入,而混合物中的环氧树脂 E-51 与固化剂顺酐按质量比 100:49.2 加入。搅拌混匀后再倾入已涂有脱模剂的预热模具中。按程序升温固化,固化完后冷却至室温。

1.2.2 浇铸体固化制度的确定

采用德国 NETZSCH 公司产 DSC 200 F3 型差示扫描量热仪对样品的放热情况进行动态 DSC 监测,确定固化工艺。用量为 5~10mg,升温速率 $\beta = dT/dt$ 分别为 5、10、15、20℃/min。

1.2.3 力学性能测试:

按照 GB/T 2570-1995 测试弯曲性能;按照 GB/T 2568-1995 测试拉伸性能。

1.2.4 玻璃化转变温度(T_g)

采用 DSC 200 F3 型差示扫描量热仪对样品的放热情况进行动态的 DSC 监测,用量为 5~10mg,对树脂固化物按 10℃/min 的升温速率进行 DSC 扫描,测试固化物的玻璃化转变温度。

2 结果与讨论

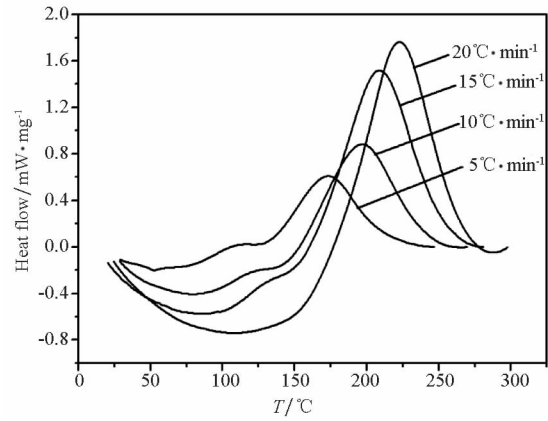
2.1 DSC 分析

采用 DSC 200 F3 型差示扫描量热仪对不同质量比 E-51 改性的 CDR-0122 样品的放热情况进行动态的 DSC 监测^[3-6],升温速率 $\beta = dT/dt$ 分别为 5、10、15、20℃/min。在选定的升温速率下,不同质量分数 E-51 改性的动态 DSC 曲线如图 1。

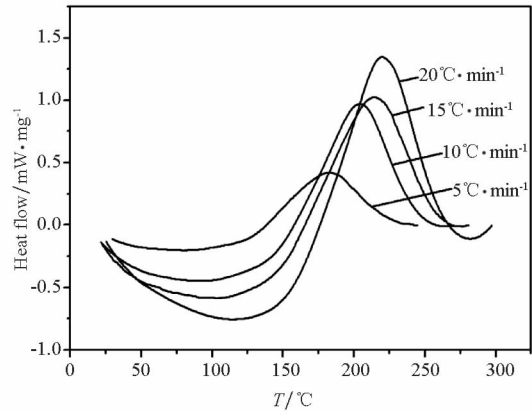
由图 1 可知,每个波峰的起始温度 T_i 、峰值温度 T_p 及终止温度 T_f 的值。同一配比的树脂体系固化反应的起始温度 T_i 、峰值温度 T_p 及终止温度 T_f 均随升温速率的增大而向高温方向移动。而不同配比的树脂体系在同一升温速率下随着 E-51 量的增加,反应的放热峰相差较大,但起始温度 T_i 、峰值温度 T_p 和终止温度 T_f 相差不大。当 CDR-0122 树脂未改性(E-51 为 0)时,升温速率为 5、10、15℃/min 时的峰形是双峰,峰形随升温速率 β 的加快而逐渐变小,定义第 1 个波峰为 μ_1 ,第 2 个波峰为 μ_2 ;而当升温速率为 20℃/min 时的峰形为单峰。这是由于当升温速率过快时,第 1 步和第 2 步反应趋于同时进行,使两个峰的区别显现得不是特别明显,因此未改性时的树脂体系可以认为具有两个放热峰,是两步反应。它的反应机理是首先醇与酸酐反应生成含酯键的羧酸;然后羧酸与环氧基开环反应,生成含羟基的中

间化合物;中间化合物中的羟基既可以再与酸酐反应生成含酯键的羧酸化合物,并且两者均可与环氧基反应。而加入 E-51 后均呈单峰。这是因为未用 E-51 改性时,CDR-0122 环氧树脂不含羟基,它直接与酸酐的反应速度非常慢,它的反

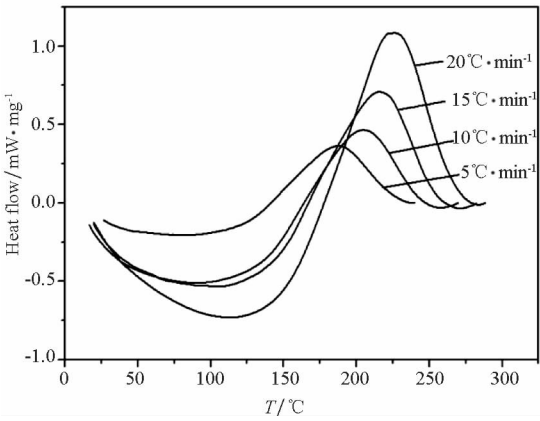
应机理是先由多元醇与酸酐反应后,再与环氧基反应。而通过 E-51 改性后,由于 E-51 与 CDR-0122 的相容性很好,环氧基和羟基赋予了其相互反应的活性,使反应一步完成。



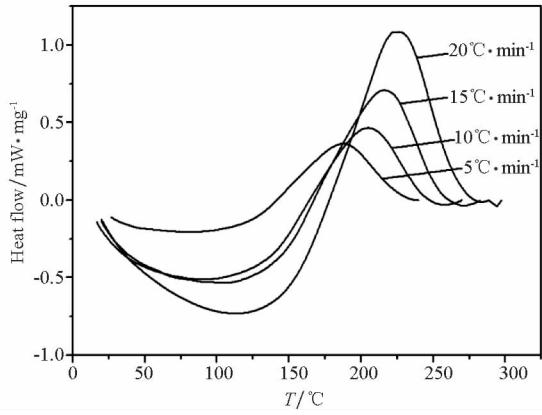
(a) 0% E-51



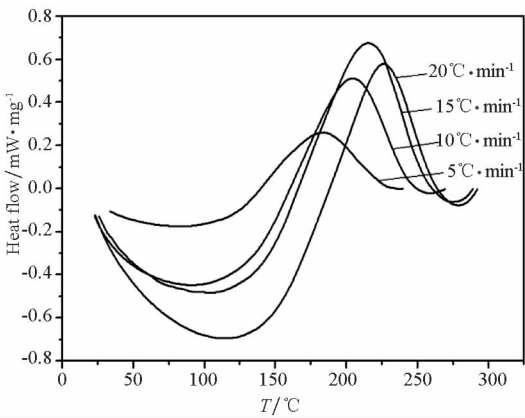
(b) 10% E-51



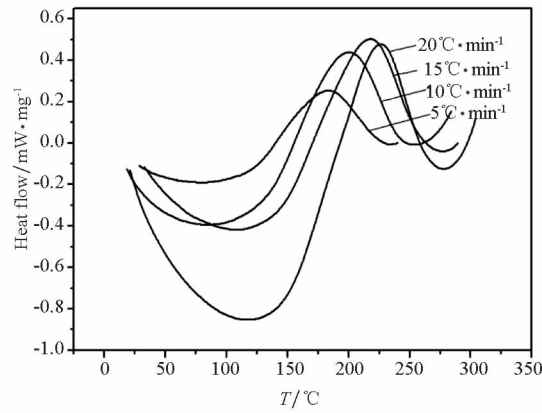
(c) 20% E-51



(d) 30% E-51



(e) 40% E-51



(f) 50% E-51

图 1 不同质量分数 E-51 改性 CDR-0122 树脂体系在不同升温速率下的动态 DSC 曲线

Fig. 1 Dynamic DSC curves of CDR-0122 of modified resin system with different mass fraction of E-51 resin

2.1.1 活化能的确定

活化能是表征体系聚合反应进行难易程度的量度。通常化学反应的活化能 E_a 大致在 40 ~

400 $\text{kJ} \cdot \text{mol}^{-1}$, 一般来说, $E_a < 40 \text{ kJ} \cdot \text{mol}^{-1}$, 则反应在室温以下即可以瞬时完成, 若 $E_a > 100 \text{ kJ} \cdot \text{mol}^{-1}$, 则要适当地进行加热, 反应才能进行。

E_a 越大,能迅速完成反应要求的温度也越高^[3]。

本实验通过 DSC 热分析数据来计算固化反应动力学参数,用 DSC 法研究固化动力学有多种计算动力学参数的方法,其中比较常用的是 Kissinger 方法。对于环氧树脂的动态固化反应 Kissinger 方程为

$$\ln \frac{\beta}{T_p^2} = \frac{E_a}{R} \cdot \frac{1}{T_p} + \ln \frac{AR}{E_a}$$

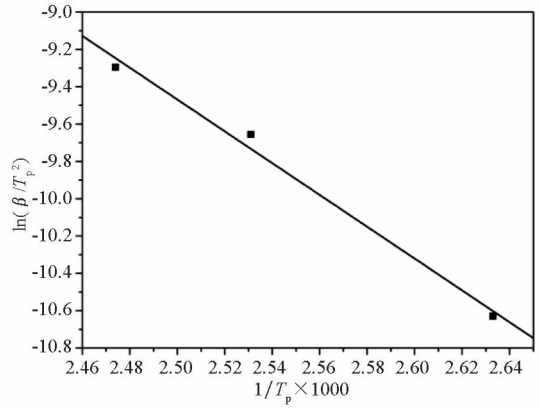
(1)

式中 β 为等速升温速率 ($K \cdot \min^{-1}$), R 为气体常数 $8.314J/mol$, T_p 为峰值温度, A 和 E_a 分别为表观频率因子和反应活化能。

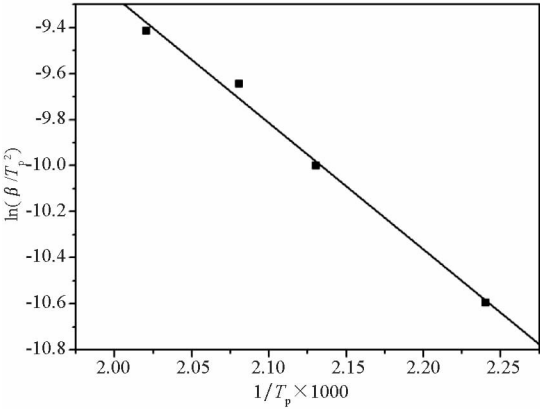
由式(1)可知频率因子:

$$A = \frac{\beta E_a \exp(E_a/RT_p)}{RT_p^2}$$

(2)



(a) μ_1



(b) μ_2

图2 未改性树脂体系 $\ln(\beta/T_p^2)$ 与 $1/T_p$ 的关系曲线

Fig. 2 The relationship between $\ln(\beta/T_p^2)$ and $1/T_p$ of unmodified resin system

由上式(1)易知 $\ln(\beta/T_p^2)$ 与 $1/T_p$ 存在线性关系,将 5、10、15、20 $^{\circ}C/min$ 升温速率下的值进行线性拟合,见图 2 和 3。由式(1)和图 2、3 可知直线的斜率即为 $-E_a/R$,则易求得不同量 E-51 改性体系的活化能见表 2 和表 3。改性前 μ_1 的活

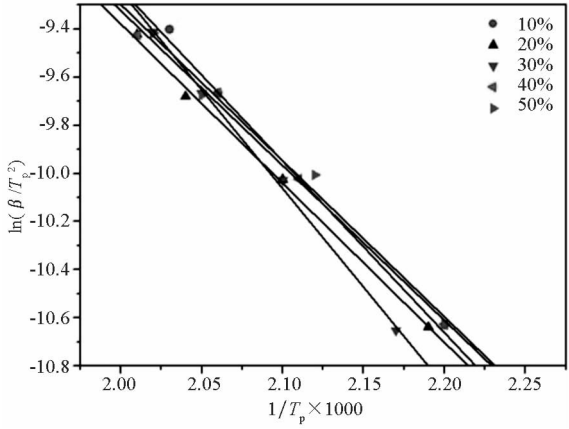


图3 不同质量分数 E-51 改性树脂体系 $\ln(\beta/T_p^2)$ 与 $1/T_p$ 的关系曲线

Fig. 3 The relationship between $\ln(\beta/T_p^2)$ and $1/T_p$ of modified resin system with different mass fraction of E-51

化能为 $60.376\text{ kJ} \cdot \text{mol}^{-1}$, μ_2 的活化能为 $45.652\text{ kJ} \cdot \text{mol}^{-1}$; E-51 改性后的活化能相比未改性时略有下降。这是由于 E-51 含有一定量的羟基,在一定程度上赋予了树脂体系相互反应的活性。由所求得的结果可知活化能较大,所以该体系需要高温固化。

表2 未改性树脂体系的反应活化能 E_a

Tab. 2 Activation energy (E_a) of unmodified resin system

	μ_1	μ_2
$E_{a1}/\text{kJ} \cdot \text{mol}^{-1}$	60.376	45.652

表3 改性后树脂体系的活化能 E_a

Tab. 3 Activation energy (E_a) of modified resin system

$m_{E-51}/m_{CDR-0122}$	10/100	20/100	30/100	40/100	50/100
$E_a/\text{kJ} \cdot \text{mol}^{-1}$	59.809	54.997	60.108	53.484	53.925

2.1.2 反应级数的确定

用 Crane 经验方程可以求出固化反应级数, Crane 方程表示为

$$\frac{d \ln(\beta)}{d(1/T_p)} = - \left(\frac{E_a}{nR} + 2T_p \right)$$

(3)

式中 β 为等速升温速率 ($^{\circ}C \cdot \min^{-1}$), R 为气体常数 $8.314J/mol$, T_p 为峰值温度, n 和 E_a 分别为反应级数和反应活化能。

当 $E_a/nR \gg 2T_p$ 时,可将 $2T_p$ 省略,因而可以得到下式:

$$\frac{d \ln(\beta)}{d(1/T_p)} \approx - \frac{E_a}{nR}$$

(4)

即 $\ln\beta$ 与 $1/T_p$ 之间存在一定的线性关系,将

5、10、15、20 ℃ /min 升温速率下的值进行线性拟合如图 4 和图 5。由图可知, $\ln\beta$ 与 $1/T_p$ 的线性关系显著, 则拟合直线的斜率为 $-E_a/nR$, 由所求得的活化能可求出不同量 E-51 改性的树脂体系的固化反应级数 n , 结果见表 4 和表 5。由结果可知反应级数相差都不大, 均趋于 1 级反应。

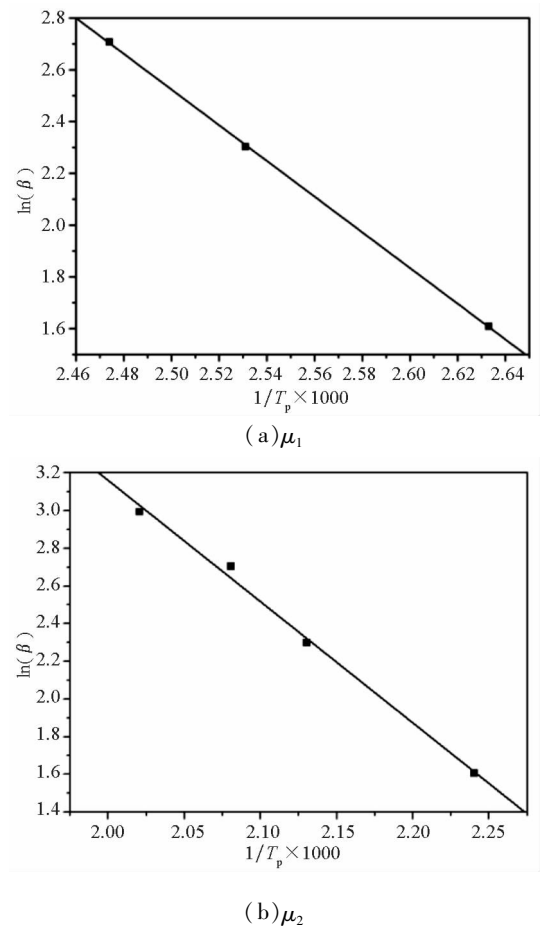


图 4 未改性树脂体系 $\ln(\beta)$ 与 $1/T_p$ 的关系曲线
Fig. 4 The relationship between $\ln(\beta)$ and $1/T_p$ of unmodified resin system

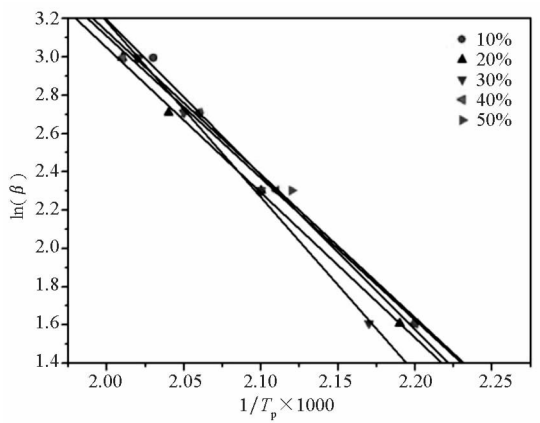


图 5 不同质量分数 E-51 改性树脂体系 $\ln(\beta)$ 与 $1/T_p$ 的关系曲线

Fig. 5 The relationship between $\ln(\beta)$ and $1/T_p$ of modified resin system with different mass fraction of E-51

表 4 未改性树脂体系的反应级数 n				
Tab. 4 Reaction order (n) of unmodified resin system				
	μ_1	μ_2		
n	0. 903	0. 853		

表 5 改性后树脂体系的反应级数 n					
Tab. 5 Reaction order (n) of modified resin system					
$m_{E-51}/$ $m_{CDR-0122}$	10/100	20/100	30/100	40/100	50/100
n	0. 905	0. 874	0. 894	0. 868	0. 868

2. 1. 3 固化制度的确定

最佳固化工艺的外推法是对固化体系进行多次不同升温速率的扫描, 分别以每个放热峰的起始温度 T_i 、峰值温度 T_p 和终止温度 T_c 为纵坐标, 以升温速率 β 为横坐标进行线性拟合, 固化反应温度 (T) 与固化升温速率 (β) 呈线性关系, 通过 $T-\beta$ 外延法得到该直线与 Y 轴的交点, 该交点的纵坐标即是外推出的升温速率 $\beta=0$ 时的起始温度、峰值温度和终止温度。运用外推法得到体系的固化制度见表 6。

表 6 改性前后树脂体系的固化制度	
Tab. 6 Solidify system of unmodified and modified resin system	
$m_{E-51}/$ $m_{CDR-0122}$	Solidify system
0/100	79. 27℃ + 103. 80℃ + 121. 66℃ + 143. 61℃ + 169. 94℃ + 207. 94℃
10/100	126. 84℃ + 180. 65℃ + 219. 91℃
20/100	123. 24℃ + 180. 01℃ + 220. 08℃
30/100	120℃ + 177. 13℃ + 221. 36℃
40/100	115. 2℃ + 169. 9℃ + 212. 6℃
50/100	118. 65℃ + 176. 82℃ + 220. 41℃

2. 2 浇铸体基本性能分析

按照表 6 的固化制度对 CDR-0122 改性前后的预聚体进行固化, 然后测其浇铸体的拉伸性能、弯曲性能和固化产物的玻璃化转变温度^[7-11]。结果见表 7。其中优选树脂浇铸体的拉伸应力-应变曲线和三点弯曲载荷下的浇铸体载荷-位移曲线见图 6 和图 7。

表7 E-51 含量对 CDR-0122 环氧树脂基本性能的影响
Tab.7 Influences of the mass fraction of E-51 on the properties of CDR-0122 epoxy resin

$m_{E-51}/m_{CDR-0122}$	Bending strength/MPa	Bending modulus/GPa	Tensile strength/MPa	Tensile modulus/GPa	$T_g/^\circ\text{C}$
0/100	77.23	4.16	36.77	3.39	253.90
10/100	110.98	3.28	38.91	3.21	251.89
20/100	99.12	3.69	52.11	3.14	248.38
30/100	117.68	3.41	56.91	3.11	246.98
40/100	113.27	2.90	46.83	2.98	234.74
50/100	114.83	3.53	46.14	3.02	197.08

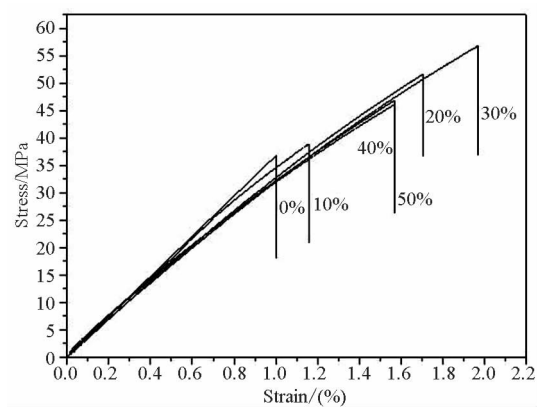


图6 优选浇铸体的拉伸应力-应变曲线
Fig.6 Stress-strain curves of resin casts

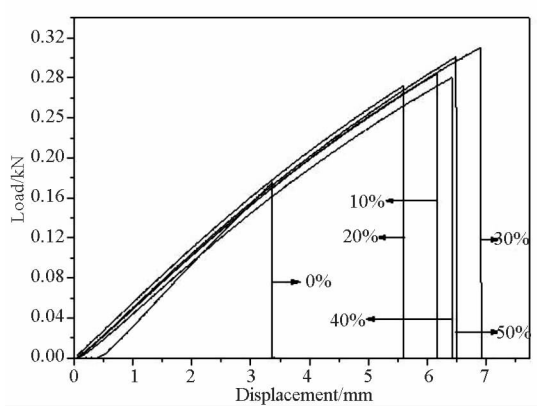


图7 三点弯曲载荷下的优选浇铸体载荷-位移曲线
Fig.7 Load-displacement curves of resin casts under three point bending

由表7、图6和图7可知,当加入质量分数30% E-51时,树脂体系的拉伸和弯曲性能最好,弯曲强度和拉伸强度分别提高了52.38%和54.77%,而弯曲模量和拉伸模量均下降。由图6可知CDR-0122树脂浇铸体的应力-应变曲线近乎为一直线,其断裂为脆性断裂;而E-51改性CDR-0122后的树脂浇铸体拉伸应力-应变曲线显示出一定的非线性,是塑性断裂,表明E-51改性后的树脂浇铸体韧性有所提高。由图7可知CDR-0122树脂浇铸体载荷-位移曲线也

近乎为一条直线,表明弯曲性能较差,断裂延伸率低,为一种高脆性基体,需要适当改性。而添加一定量E-51后的CDR-0122树脂体系,浇铸体的载荷-位移曲线呈现弧度状,弯曲强度、弯曲极限绕度都明显提高,表明E-51改性后的树脂浇铸体韧性有所提高。加入少量的E-51改性时玻璃化转变温度下降并不明显,但当E-51的加入量超过30%时玻璃化转变温度明显下降。这说明一定量的E-51树脂改性CDR-0122后不仅具有良好的耐热性,且具备较好的力学性能。这主要是由于CDR-0122本身的刚性较强,随着改性树脂E-51的加入,增加了交联点间的分子量,形成了刚性和柔性链接并存的交联网络结构。当受到外力作用时,由于CDR-0122体系的脆性得到了改善,抵抗开裂的能力增强,因此体系呈现较高的力学性能。

3 结 论

- (1)未改性的树脂体系有两个波峰,分两步进行反应;改性后的树脂体系只有一个波峰,反应一步完成。当加入30% E-51时树脂体系的活化能为68.108 kJ·mol⁻¹,反应级数为0.894。树脂体系的活化能较大,适合高温固化。它的固化制度可定为120℃+177.13℃+221.36℃。
- (2)当用30% E-51改性时,树脂体系固化物的力学性能最好,弯曲强度和拉伸强度分别为117.68 MPa和56.91 MPa,相比未改性时分别提高了52.38%和54.77%,但玻璃化转变温度相比改性前也下降了4.8℃,为255.01℃。综合考虑各方面的性能,确定当加入E-51的质量分数为30%时,CDR-0122的树脂体系最好,不仅固化物的力学性能较高,而且耐温性能也较好。

参考文献 (References)

[1] Wani M F, Gandhi O P. Development of maintainability index for mechanical system[J]. Reliability Engineering and System Safety, 1999, 65: 259 – 270.

[2] 黎放, 杨元, 杨建军. 装备维修性设计参数综合评估方法[J]. 工程设计学报. 2008, 15(1): 11 – 16.

LI Fang, YANG Yuan, YANG Jianjun. Integrative evaluation on parameters of maintainability design of mechanical equipment[J]. Journal of Engineering Design, 2008, 15(1): 11 – 16. (in Chinese)

[3] Slavila C A, Decreuse C, Ferney M. Fuzzy approach for maintainability evaluation in the design process [J]. Concurrent Engineering, 2005, 13: 291 – 299.

[4] Chen L, Cai J G. Using vector projection method to evaluate maintainability of mechanical system in design review [J]. Reliability Engineering and System Safety, 2003, 81: 147 – 154.

[5] 国防科学技术工业委员会. GJB/Z91—1997 维修性设计手册[S]. 北京: 中国标准出版社, 1997.

The committee of National Defense Scientific and Technical Industry. GJB/Z91 – 1997 maintainability design manual[S]. Beijing: Standards Press of China, 1997. (in Chinese)

[6] 甘茂治, 吴真真. 维修性设计与验证[M]. 北京: 国防工业出版社, 1995.

GAN Maozhi, WU Zhenzhen. Maintainability design an evaluation[M]. Beijing: Defense Industry Press, 1995. (in Chinese)

[7] 赵克勤. 集对分析及其初步应用[M]. 杭州: 浙江科学技术出版社, 2000.

ZHAO Keqin. Set pair analysis and its preliminary applications[M]. Hangzhou: Zhejiang Science and Technology Press, 2000. (in Chinese)

[8] 杨俊杰, 周建中, 方仍存, 等. 基于集对分析的不确定多属性决策方法[J]. 控制与决策. 2008, 23(12): 1423 – 1426.

YANG Junjie, ZHOU Jianzhong, FANG Rengcun, et al. Uncertain multi-attribute decision making methods based on set pair analysis[J]. Control and Decision, 2008, 23(12): 1423 – 1426. (in Chinese)

[9] Chiou H K., Tzeng G H. Fuzzy multiple-criteria decision-making approach for industrial green engineering [J]. Environmental Management. 2002, 30(6): 816 – 830.

[10] 刘英平, 杨素君, 陈晓霞. 不确定信息的机械产品设计方案评价方法[J]. 机械科学与技术. 2008, 27(7): 946 – 949.

LIU Yingping, YANG Sujun, CHEN Xiaoxia. Study of evaluation method for mechanical product design schemes with uncertainty information [J]. Mechanical Science and Technology for Aerospace Engineering. 2008, 27(7): 946 – 949. (in Chinese)

(上接第 43 页)

参考文献 (References)

[1] Mohammad M, Fangming D, Naiffer R, et al. Increased flexural modulus and strength in SWNT/epoxy composites by a new fabrication method [J]. Polymer, 2006, 47 (1): 293 – 298.

[2] Varley R J, Hodgkin J H, Hawthorne D G. Toughening of a trifunctional epoxy system; Part VI[J]. Polymer, 2001, 42: 3847 – 3858.

[3] 李晓丹, 王维, 郭淑齐, 等. 耐高温树脂的固化动力学分析及其力学性能[J]. 纤维复合材料, 2009, 24(4): 24 – 26.

LI Xiaodan, WANG Wei, GUO Shuqi, et al. The analysis of curing kinetics of heat resistant resin and research of its composites performances[J]. Fiber Composites, 2009, 24(4): 24 – 26. (in Chinese)

[4] 代晓青, 肖加余. 等温 DSC 法研究 RFI 用环氧树脂固化动力学研究 [J]. 复合材料学报, 2008, 25 (4): 18 – 23. (in Chinese)

DAI Xiaqing. XIAO Jiayu. Curing kinetics of epoxy resin for RFI process using isothermal DSC [J]. Acta Materiate Compositae Sinica, 2008, 25(4): 18 – 23. (in Chinese)

[5] Barton J M. The application of differential scanning calorimetry (DSC) to the study of epoxy resin curing reactions [J]. Advances in Polymer Science, 1985, 72: 1 – 11.

[6] 谢旻, 顾轶卓, 李敏, 等. 碳纤维/双马树脂预浸料固化过

程动态力学性能[J]. 复合材料学报, 2010, 27(4): 52 – 58.

XIE Min, GU Yizhuo, LI Min, et al. Dynamic mechanical analysis of the curing of carbon fiber/bismaleimide resin prepreg [J]. Acta Materiate Compositae Sinica, 2010, 27 (4): 52 – 58. (in Chinese)

[7] Yun N G, Won Y W, Kim S C. Toughening of carbon fiber/epoxy composite by inserting polysulfone film to form morphology spectrum[J]. Polymer, 2004, 45(20): 6953 – 6958.

[8] Mimura K, Ito H, Fujioka H, Improvement of thermal and mechanical properties by control of morphologies in PES modified epoxy resins[J]. Polymer, 2000, 41 (12): 4451 – 4459.

[9] Catalani A, Bonicelli M G. Kinetics of the curing reaction of a diglycidyl ether of bisphenol A with a modified polyamine[J]. Thermochimica Acta, 2005, 438(1): 126 – 129.

[10] Li P, Yang X P, Yu Y H, et al. Cure kinetics, microheterogeneity, and mechanical properties of the high-temperature cure of vinyl ester resins[J]. Journal of Applied Polymer Science, 2003, 92(2): 1124 – 1133.

[11] Kamal M R, Sourour S. Kinetics and thermal characterization of thermoset cure [J]. Polymeric Materials Science and Engineering, 1973, 13: 59 – 64.

非均布应力场中内埋裂纹的应力强度因子*

袁杰红¹, 孙鹏飞¹, 段静波²

(1. 国防科技大学 指挥军官基础教育学院, 湖南 长沙 410072;
2. 国防科技大学 航天与材料工程学院, 湖南 长沙 410073)

摘要: 在内埋裂纹线性线弹簧模型的基础上, 通过引入二维权函数将裂纹面上的非均布载荷进行均布化等效, 求解了中心内埋椭圆形裂纹在沿板厚非均匀分布应力场中的应力强度因子, 列出了问题的奇异积分方程, 利用 Gauss-Chebyshev 方法求解了在 4 种应力场分布情形下的数值结果, 并与已有文献的解进行了比较, 当 $a_0/c_0 < 0.4$ 、 $a_0/h \leq 0.3$ 时, 两者结果具有较好的一致性, 表明了本文方法的合理性和可靠性。

关键词: 线弹簧模型; 中心内埋裂纹; 非均布应力场; 权函数; 应力强度因子
中图分类号: O346.1 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0044-04

The stress intensity factor of embedded cracks
in non-uniform stress fields

YUAN Jiehong¹, SUN Pengfei¹, DUAN Jingbo²

(1. College of Basic Education for Commanding Officers, National University of Defense Technology, Changsha 410072, China;
2. College of Aerospace and Materials Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The stress intensity factor of a center embedded elliptical crack in non-uniform stress fields, which is along the thickness direction of the plate, is gained based on the linear line-spring model for embedded cracks. The two dimensional weight function is used to transform the non-uniform stress field to an equivalent uniform one. The singular integral equations are formulated and the numerical results in four cases of stress distributions are gained by Gauss-Chebyshev method. The results are in good accordance with those given in the previous literature when $a_0/c_0 < 0.4$, $a_0/h \leq 0.3$, and the rationality and reliability of this method are demonstrated.

Key words: line-spring model; center embedded crack; non-uniform stress field; weight function; stress intensity factor

内埋裂纹(图 1)是一种常见的未穿透裂纹, 如结构中的气孔、夹渣等缺陷都可能发展为内埋裂纹, 并且其裂纹面常受到残余应力等复杂载荷的作用, 应力场的非均布性使得内埋裂纹应力强度因子的求解更加困难。对于非均布应力场中的裂纹问题, 有限元法和权函数法是目前比较有效的手段。计算量大、费用高和准备工作复杂等问题限制了有限元法的应用推广, 而对于权函数法, 虽然有一系列的报道^[1-3], 但是三维裂纹问题的权函数求解复杂, 不利于工程实际应用。

线弹簧模型最初由 Rice 等^[4]提出, 用于计算表面裂纹应力强度因子, 袁杰红等^[5]将其推广至内埋裂纹问题的求解, 建立了在远场作用拉伸和弯曲载荷时无限平板内埋裂纹的线性线弹簧模型, 但是对于裂纹面承受非均布载荷的情况, 则不能直接应用该模型求解。鉴于权函数法求解非均布应力场中裂纹问题的独特优势, 如果将线弹簧模型和权函数法相结合, 则有望求解非均布应力

场中的内埋裂纹问题。本文在无限平板内埋裂纹线性线弹簧模型^[5]的基础上, 通过引入二维权函数将沿板厚非均匀分布的应力场进行等效处理, 求解了中心内埋椭圆形裂纹在沿板厚非均匀分布应力场中的应力强度因子, 给出了相应的数值结果, 并通过与文献[1]的比较, 验证了本文方法的正确性。

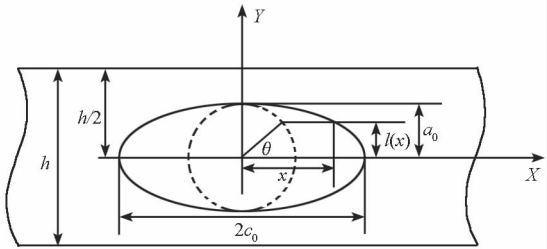


图 1 中心内埋裂纹示意图
Fig. 1 A center embedded elliptical crack

* 收稿日期: 2011-06-09
作者简介: 袁杰红(1965—), 男, 湖南长沙人, 教授, 博士, 硕士生导师, E-mail: yjh210048@163.com

1 无限平板内埋裂纹线性线弹簧模型描述

如图 1 和图 2(a) 所示,在无穷远处单位宽度上作用有外力 N^∞ 和外力矩 M^∞ 的内埋椭圆形裂纹(a_0 、 c_0 分别为短轴和长轴半长)的无限平板,可以转化为裂纹面上作用有线弹簧的长度为 $2c_0$ 的穿透裂纹平板(图 2(b))。线弹簧的本构关系可由相应位置的平面应变边裂纹板条所受的广义力和由于裂纹存在引起的附加广义位移的关系来确定(图 2(c));内埋裂纹前缘各点的应力强度因子等于相应位置的内裂纹板条的应力强度因子。这样,求解无限平板内埋裂纹问题便转化为求解穿透裂纹平板问题和一个作用有弹簧 $N(x)$ 、 $M(x)$ 力的内裂纹板条问题,它们都是二维问题。由前者可获得平板性能方程,由后者可得到线弹簧本构关系表达式,两者联立后,可得到内埋裂纹问题的解。

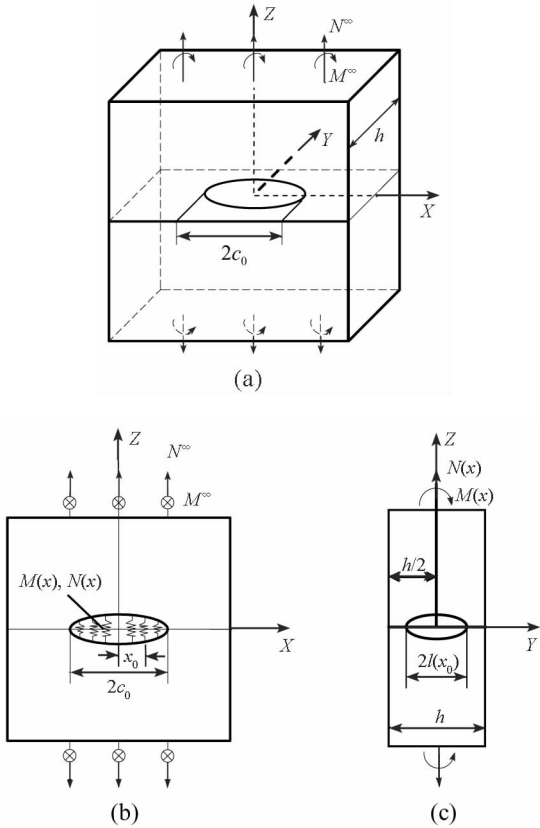


图 2 内埋裂纹线弹簧模型

Fig. 2 The line-spring model of embedded cracks

2 非均匀应力场的处理

对于在远场作用载荷 N^∞ 和 M^∞ ,在裂纹面作用非均匀载荷 $\sigma(\bar{x},\bar{y})$ ($\bar{x} = \frac{x}{c_0}, \bar{y} = \frac{y}{h}$ 均为无量纲坐标, h 为板厚)的中心内埋裂纹,上述线弹簧

模型由于非均匀载荷的存在而不能直接应用,但由上节的描述可知,内埋裂纹前缘各点的应力强度因子等于相应位置的内裂纹板条的应力强度因子。那么,可以取在 $\bar{x} = \bar{x}_0$ 处(图 2(b)), $|\bar{x}_0| \leq 1$, $\bar{x} = x/c_0, \bar{x}_0 = x_0/c_0$) 平行于 YOZ 面的一内裂纹板条(图 2(c)) 进行研究。在线弹性范围内,根据叠加原理可将内裂纹板条的应力强度因子分为两部分求解:一部分是由远场载荷 N^∞ 和 M^∞ 引起的;另一部分是由非均匀载荷 $\sigma(\bar{x}_0,\bar{y})$ 引起的。

如图 3(a) 所示为在裂纹面上作用有关于 Y 轴和 Z 轴对称但非均匀分布应力 $\sigma(\bar{x}_0,\bar{y})$ 的内裂纹板条,根据权函数理论,当 $\frac{l(\bar{x}_0)}{h} \leq 0.3$ 时,其应力强度因子可由下式^[6] 获得:

$$K_I = f\sigma_0 \sqrt{\pi l(\bar{x}_0)} \tag{1}$$

其中, $f = \int_0^\xi \frac{\sigma(\bar{x}_0,\bar{y})}{\sigma_0} \sqrt{\frac{h}{\pi \xi}} m(\bar{y},\xi) d\bar{y}, m(\bar{y},\xi) = \frac{2}{\sqrt{\pi \xi h}} \left(\frac{\xi}{\sqrt{\xi^2 - \bar{y}^2}} + \frac{\pi}{2} \sqrt{\xi^2 - \bar{y}^2} \tan \pi \xi \right)$ 为权函数, σ_0 为原点处应力值, $\xi = \frac{l(\bar{x}_0)}{h}, l(\bar{x}_0) = a_0 \sqrt{1 - (\bar{x}_0)^2}, \bar{y} = y/h, h$ 为板厚。

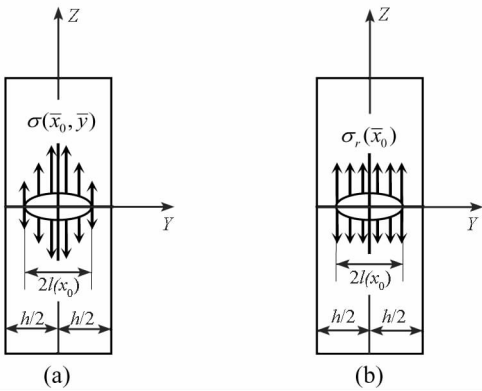


图 3 内裂纹板条示意图

Fig. 3 The slice with a center embedded crack

对于作用均匀应力场 $\sigma_r(\bar{x}_0)$ 的内裂纹板条(图 3(b)),其应力强度因子的表达式为

$$K_I = \sigma_r(\bar{x}_0) \sqrt{\pi l(\bar{x}_0)} \cdot \sqrt{\sec \frac{\pi l(\bar{x}_0)}{h}} \tag{2}$$

如果 $\sigma_r(\bar{x}_0)$ 和 $\sigma(\bar{x}_0,\bar{y})$ 两者引起的应力强度因子相等,那么 $\sigma_r(\bar{x}_0)$ 就可以作为非均匀应力场 $\sigma(\bar{x}_0,\bar{y})$ 的等效均匀应力场。令式(1)和式(2)相等,可得等效关系式为

$$\sigma_r(\bar{x}_0) = f\sigma_0 \sqrt{\cos \pi \xi} \tag{3}$$

其中, f, σ_0 和 ξ 均如前所指。这样处理后,由 $\sigma(\bar{x}_0,\bar{y})$ 引起的应力强度因子便可等效为由均匀应力场 $\sigma_r(\bar{x}_0)$ 引起。

由于 x_0 具有一般性,那么当 $|x| \leq 1$ 时,式 (3) 可以写成

$$\sigma_r(\bar{x}) = f\sigma_0 \sqrt{\cos\pi\xi} \tag{4}$$

其中 $f = \int_0^\xi \frac{\sigma(\bar{x},\bar{y})}{\sigma_0} \sqrt{\frac{h}{\pi\xi}} m(\bar{y},\xi) d\bar{y}, m(\bar{y},\xi)$ 如前所指, $\xi = \frac{l(\bar{x})}{h}, l(\bar{x}) = a_0 \sqrt{1 - \bar{x}^2}$ 。

3 非均布应力场中内埋裂纹的应力强度因子

将非均布应力场 $\sigma(\bar{x},\bar{y})$ 等效为均布应力场 $\sigma_r(\bar{x})$ 后,根据 Bueckner 等效原则, $\sigma_r(\bar{x})$ 便可以和远场作用的载荷 N^∞ 和 M^∞ 进行叠加,根据文献 [5] 可直接写出求解非均布应力场中内埋裂纹应力强度因子的无量纲形式的奇异积分方程

$$\int_{-1}^1 \begin{bmatrix} \frac{1}{t-\bar{x}} + k_{11}(\bar{x},\bar{t}) & k_{12}(\bar{x},\bar{t}) \\ k_{21}(\bar{x},\bar{t}) & \frac{3}{\bar{t}-\bar{x}} + 3k(\bar{x},\bar{t}) + k_{22}(\bar{x},\bar{t}) \end{bmatrix} \begin{Bmatrix} \mu_1(\bar{t}) \\ \mu_2(\bar{t}) \end{Bmatrix} d\bar{t} = -\frac{2\pi}{E} \begin{Bmatrix} \sigma_M^\infty + \sigma_r(\bar{x}) \\ \frac{c_0}{h} \sigma_B^\infty \end{Bmatrix} \tag{5}$$

其中, $k(\bar{x},\bar{t}), k_{i,j}(\bar{x},\bar{t}), i,j = 1,2$ 为已知核函数, E 为弹性模量, c_0 为裂纹长轴半长, h 为板厚, $\mu_1(\bar{t}), \mu_2(\bar{t})$ 为未知位错密度函数, $\bar{t} = t/c_0, \bar{x} = x/c_0, \sigma_M^\infty = N^\infty/h, \sigma_B^\infty = 6M^\infty/h^2, \sigma_r(\bar{x}) = f\sigma_0 \sqrt{\cos\pi\xi}$ 为等效均布应力场, 其他均如前所指。

利用 Gauss-Chebyshev 方法^[7] 求解方程 (5) 的基本未知量 $\mu_1(t)$ 和 $\mu_2(t)$ 。之后,通过平板性能方程^[5] 和内埋裂纹板条应力强度因子的表达式^[5] 就可求得坐标为 $\bar{x}_k(\bar{x}_k$ 为 Chebyshev 多项式

零点)点处的线弹簧内力和裂纹前缘对应点的应力强度因子。为能得到裂纹前缘任意点的应力强度因子,可对所得对应 \bar{x}_k 的应力强度因子作多项式拟合。

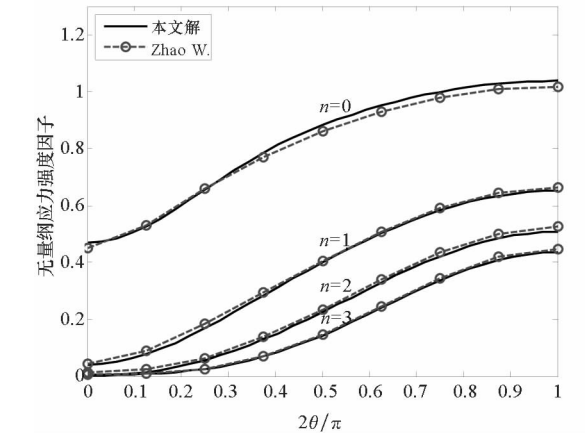
4 数值结果及分析

令远场拉伸力和弯矩为零,并假设非均布载荷的分布规律为 $\sigma(\bar{x},\bar{y}) = \sigma_0 \left(\frac{\bar{y}}{a_0} \right)^n$ (σ_0 为原点处应力值, $\bar{y} = \frac{y}{h}, \bar{a}_0 = \frac{a_0}{h}$)。本文分别计算了在 $n = 0, 1, 2, 3$ 四种情形下,不同 a_0/c_0 及 a_0/h 的中心内埋裂纹的应力强度因子,为了与文献 [1] 的解进行对比,把对应应力强度因子 K_I 无量纲化:

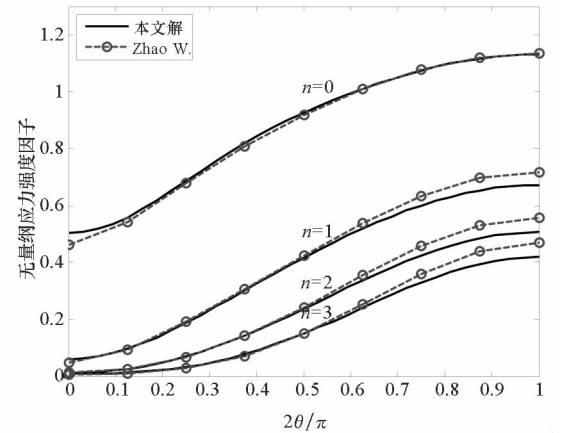
$$\bar{K}_I = \frac{K_I}{\left(\sigma_0 \sqrt{\frac{\pi a_0}{Q}} \right)} \tag{6}$$

其中, $Q = 1 + 1.464(a_0/c_0)^{1.65}$ 为裂纹形状因子。需要指出的是文献 [1] 计算的为有限宽板, c_0/b (b 为板宽) 的值取为 0.1。计算时取泊松比 $\nu = 0.3$, 计算结果及对比情况如图 4(a) ~ (f) 所示。

图 4(a) ~ 4(f) 表示在四种不同的载荷分布形式下,几种中心内埋椭圆裂纹前缘各点(以椭圆参数角 θ 表示,图 1) 应力强度因子的变化规律。可以看出,本文与文献 [1] 的结果有完全相同的变化趋势。且在 $a_0/c_0 < 0.4, a_0/h \leq 0.3$ 时,两者的结果具有较好的一致性;随着 a_0/c_0 的增大,两者相对误差有所增大,这是各线弹簧之间由于裂纹前缘曲率引起相互作用造成的,且这种相互作用随 a_0/c_0 的增大而增大,而线弹簧本身并没有考虑此种相互作用。此外,文献 [1] 中有限板宽的影响也是不可忽略的因素。



(a) $a_0/c_0 = 0.2 \quad a_0/h = 0.1$



(b) $a_0/c_0 = 0.2 \quad a_0/h = 0.25$

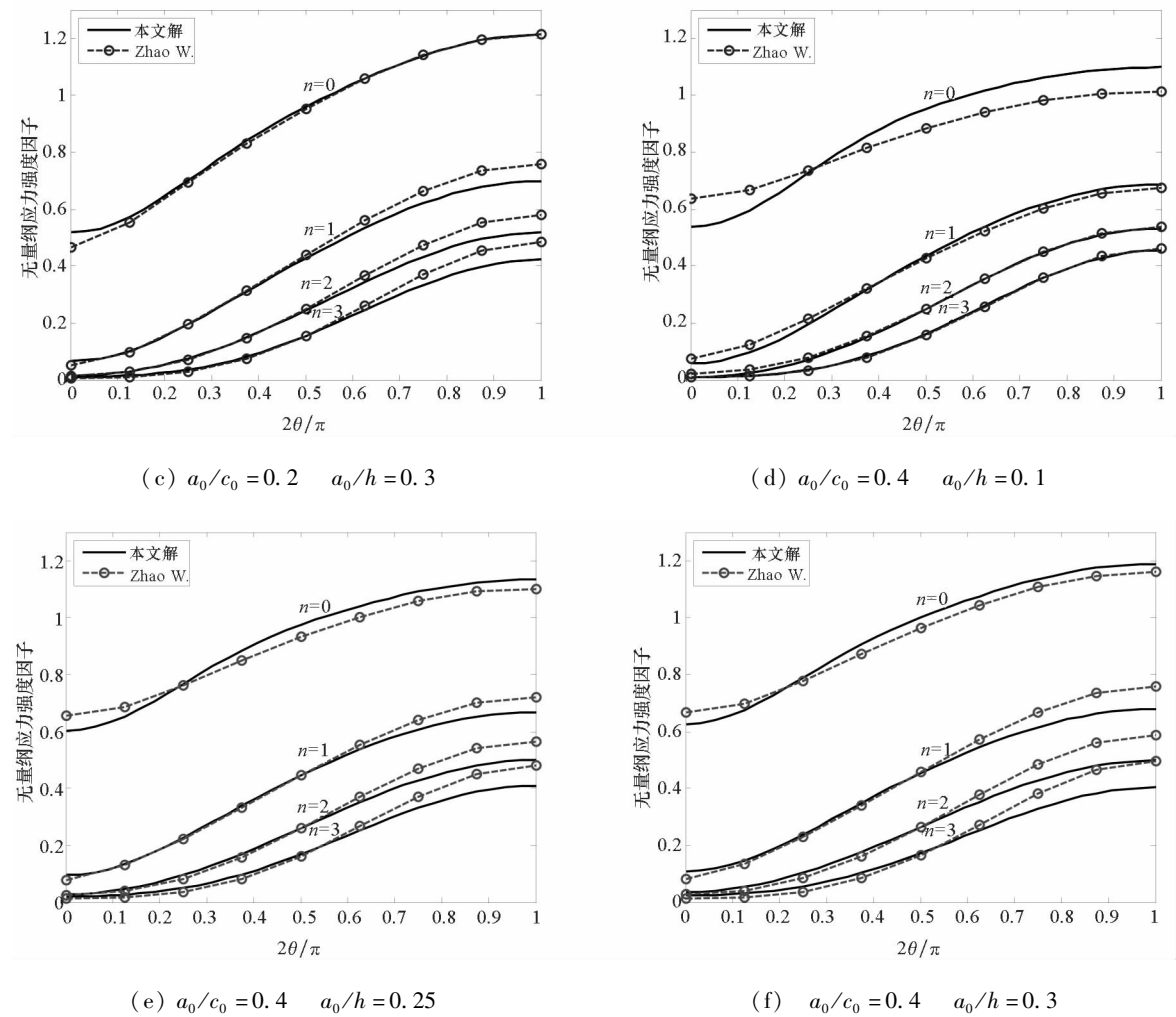


图4 中心内埋裂纹应力强度因子
Fig.4 The stress intensity factors of the center embedded crack

5 结 论

(1) 计算结果表明:本文所建立的非均匀应力场中内埋裂纹应力强度因子的求解方法是合理可靠的;

(2) 受所引用权函数适用范围的限制,本文方法的适用对象为沿板厚非均匀应力场中 $a_0/c_0 < 0.4$ 、 $a_0/h \leq 0.3$ 的中心内埋椭圆裂纹,但是这已经覆盖了工程实际中的大多数情况,且有足够的精度;

(3) 本文利用线弹簧模型将三维问题转化为二维问题,计算量小、编程方便,适合于工程实际应用。

参考文献 (References)

[1] Zhao W, Wu X R, Yan M G. Weight function method for three dimensional crack problems-I. basic formulation and application to an embedded elliptical crack in finite plates [J]. Engineering Fracture Mechanics, 1989, 34(3): 593 – 607.

[2] Wang X, Lambert S B, Glinka G. Approximate weight functions for embedded elliptical cracks [J]. Engineering Fracture Mechanics, 1998, 59: 381 – 392.

[3] Krasowsky A J, Orynyak I V, Gienko A Y. Approximate closed form weight function for an elliptical crack in an infinite body [J]. International Journal of Fracture, 1999, 99: 117 – 130.

[4] Rice J R, Levy N J. The part-through surface cracks in an elastic plate [J]. Journal of Applied Mechanics, 1972, 39: 185 – 194.

[5] 袁杰红, 唐国金, 周建平, 等. 无限平板内埋裂纹线弹簧模型[J]. 固体力学学报, 1999, 20(1): 69 – 75.

YUAN Jiehong, TANG Guojin, ZHOU Jianping, et al. The line-spring model for embedded crack in an infinite plate [J]. Acta Mechanica Solida Sinica, 1999, 20(1): 69 – 75. (in Chinese)

[6] Wu X R. Approximate weight functions for center and edge cracks in finite bodies [J]. Engineering Fracture Mechanics, 1984, 20(1): 35 – 49.

[7] Erdogan F, Gupta G D. On the numerical solution of singular integral equations [J]. Quarterly of Applied Mathematics, 1972, 29: 525 – 534.

RBF神经网络在单轴旋转惯导系统轴向陀螺漂移辨识中的应用*

于旭东¹,魏学通²,李莹²,龙兴武¹
(1. 国防科技大学 光电科学与工程学院,湖南 长沙 410073;
2. 中国人民解放军 92941 部队 95 分队,辽宁 葫芦岛 125000)

摘要:在激光陀螺单轴旋转惯性导航系统中,单轴旋转可以自动补偿垂直于旋转轴上的惯性器件误差,却不能消除旋转轴方向上惯性器件的误差,因此单轴旋转惯性导航系统的导航精度主要由轴向陀螺漂移决定。提出了一种基于径向基函数神经网络的轴向陀螺漂移辨识方法,利用系统纬度误差和温度变化量作为训练集,针对系统热态、冷态两种情况对 RBF 神经网络进行训练,对轴向陀螺漂移的辨识精度达到 0.0003°/h。试验结果表明:该方法能够有效地辨识轴向陀螺漂移,使系统达到较高的导航精度,满足实际应用的需

关键词:激光陀螺;惯导系统;单轴旋转;陀螺漂移;RBF 神经网络
中图分类号:U666.1 **文献标志码:**A **文章编号:**1001-2486(2012)03-0048-05

Application of Radial Basis Function Network for Identification of Axial RLG Drifts in Single-axis Rotation Inertial Navigation System

YU Xudong¹, WEI Xuotong², LI Ying², LONG Xingwu¹
(1. College of Opto-electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. Element 95 Unit 92941 of PLA, Huludao 125000, China)

Abstract:In the single-axis rotation inertial navigation system with ring laser gyroscope (RLG), the single-axis rotation can compensate the vertical errors of the inertial apparatus automatically, but cannot compensate the axial vertical errors, so the precision of the system is determined by the drift of the axial RLG. A novel identification method based on radial basis function network is proposed for the axial RLG drift. The inputs of the network are the latitude error and change of the temperature, and the network is trained for steady and non-steady state, in which the identification capability is less than 0.0003°/h. The experiments show that this method can estimate the axial RLG drift efficaciously, and the result of the navigation is excellent and can meet the practical demand.

Key words: ring laser gyroscope; inertial navigation system; single-axis rotation; gyro drift; RBF neural networks

激光陀螺惯性导航系统能够自主地提供载体的位置、速度和姿态等信息,是舰船安全航行的保障,也是舰载武器系统发射的姿态基准^[1-2]。随着对惯性导航精度要求的不断提高,单纯地提高惯性器件(特别是陀螺)的精度很难满足使用要求,因此必须采用系统补偿(如旋转调制方案)等关键技术^[3]。西方军事强国早在 20 世纪 80 年代初期就开始对激光陀螺应用于长时间惯性导航技术进行了研究^[4-5],现已逐步取代气浮陀螺惯导系统(AN/WSN-3)、液浮陀螺惯导系统(AN/WSN-5)和静电陀螺惯导系统(ESGN),广泛装备多种舰只和潜艇^[6]。AN/WSN-7 系列激光陀螺惯导系统已经成为美国海军水面舰船和潜艇的标准设备,已装备除三叉戟核潜艇以外的所有美海军舰艇^[7]。随着我国军事装备的日益现代化,对高精度惯性导航系统的需求也越来越迫切,因此研制高精度的惯性导航系统具有十分迫切的现实需要。

目前,旋转惯导系统的旋转方案主要有单轴旋转和双轴旋转两种方式^[8-9]。从原理上讲,单轴旋转可以消除垂直于旋转轴方向上的惯性器件误差,但不能消除旋转轴上的惯性器件误差,特别是轴向陀螺漂移,因此单轴旋转惯导系统的精度主要由轴向陀螺漂移决定^[10]。由于轴向陀螺的可观测性很低,单轴旋转惯导系统初始对准时间一般采用 16h,其中 4h 用于完成初始对准和水平惯性器件漂移的测定,12h 用于精确辨识轴向陀螺漂移。

仿真表明:单轴旋转惯导系统若要取得 1nm/72h 的定位精度,需要轴向陀螺的辨识精度

* 收稿日期:2011-09-20
作者简介:于旭东(1982—),男,吉林长春人,讲师,博士,E-mail: wind0909@163.com

优于 $0.0005^{\circ}/\text{h}$ 。文献[11]提出了一种基于姿态解算的轴向陀螺漂移的估计方法,单轴旋转惯导系统的定位精度达到 $1\text{nm}/24\text{h}$ 。本文提出了一种新颖的基于 RBF 神经网络的轴向陀螺漂移辨识方法,以系统纬度误差和温度变化量作为训练集,以系统定位误差最小为目标函数,针对系统热态、冷态两种情况对 RBF 神经网络进行训练,精确辨识轴向陀螺漂移,提高系统导航精度。

1 RBF 神经网络

在神经网络中,常用反向传播 BP 网络来实现辨识功能,但 BP 网络存在局部最优问题,且训练速度慢、效率低,而 RBF 神经网络则在一定程度上克服了这些问题,它用局部逼近的总和达到对训练数据的全局逼近,可实现全局最优^[1213]。此外,RBF 网络是一种特殊的两层网络,其参数由于 RBF 中心的固定而得以线性化,因此隐层充当了无调节参数的固定的非线性转化环节,它将输入空间映射到一个新空间,唯一的调节参数是线性叠加权值,这些参数可以采用线性最小方差 (LS) 得到,这就是 RBF 网络的优势所在。基于以上原因,本文选择 RBF 网络对轴向陀螺常值漂移进行辨识。

具有 n 个输入和一个输出的 RBF 网络结构实现了 $f_r: \mathbf{R}^n \rightarrow \mathbf{R}$ 的映射:

$$f_r(x) = \omega_0 + \sum_{i=1}^m w_i \Phi(\|x - c_i\|) \quad (1)$$

其中 $x \in \mathbf{R}^n$ 为输入量; $\Phi(\cdot)$ 为非线性转化函数, $w_i (1 \leq i \leq m)$ 为加权值 (或参数), $c_i \in \mathbf{R}^n (1 \leq i \leq m)$ 为第 i 个隶属度函数的中心, m 为中心个数。

由图 1 可知,RBF 网络输入到隐单元的权值固定为 1,所以可以看作由含有径向基神经元的隐层和含有线性神经元的输出层两层组成。隐层中每一个神经元中心和局部感受域决定了 RBF 的位置和宽度,输出层神经元将隐层的 RBF 加权叠加,只要选择足够的隐层神经元、适当的中心、局部感受域和加权值,RBF 网络可以任意精度拟合任意函数。在 RBF 网络中假设非线性转化函数 $\Phi(\cdot)$ 和 c_i 是固定的,给定一组输入和相对应的期望输出,加权值 w_i 可通过 LS 法确定。 $\Phi(\cdot)$ 的选择对网络性能的影响并不大,本文采用的基函数为高斯函数,描述为

$$\Phi(u) = \exp\left(-\frac{u^2}{\delta^2}\right) \quad (2)$$

其中 δ 为实常数,它决定了该隶属函数围绕中心

点的宽度。 $\Phi(\cdot)$ 具有局部感受的特点,即只在中心周围的一部分区域内反应较强,体现了大脑皮质层的反应特点。对网络性能影响较大的是中心 c_i 的选择,实际中心往往选择为数据的子集。

由于轴向陀螺漂移与惯导系统的输入输出存在非线性关系,采用传统的系统辨识方法很难对其进行精确的辨识,神经网络辨识不受非线性模型的限制。它依据被控系统的输入输出数据对通过学习得到一个描述系统输入输出关系的非线性映射。给定一个输入,即可得到一个输出,而不需要知道输入和输出之间存在着怎样的数学关系。

2 基于 RBF 神经网络的轴向陀螺漂移辨识

2.1 辨识时间和输入样本的确定

系统输出的经纬度误差最容易观测,因此考虑通过系统的经纬度误差信息辨识轴向陀螺漂移。图 1 给出了不同导航实验中经过 4h 初始对准和水平方向惯性器件漂移测定后 12h 内系统输出的经纬度误差曲线,从图中可以看出,在导航 12h 时,纬度误差输出为正时,经度误差时正时负,因此若采用经度误差作为观测量时,可能导致辨识的轴向陀螺漂移出现奇异解,因此只选用纬度误差作为观测量来估算轴向陀螺漂移。

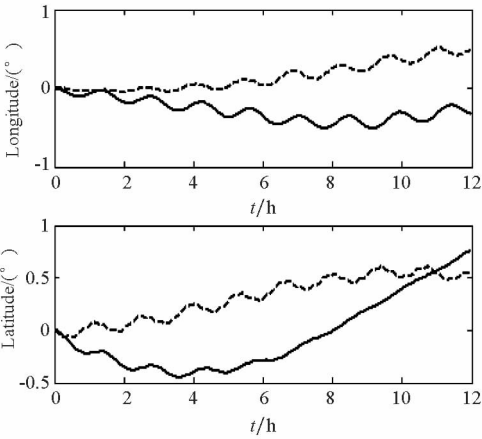


图 1 不同导航实验 12h 系统输出经纬度误差曲线
Fig. 1 Longitude and latitude error in 12 hours between different navigation tests

在静基座下,对纬度误差产生影响的主要是轴向陀螺漂移、初始对准误差、加速度计漂移等,且按照正余弦规律在系统中传播,具体形式如表 1 所示^[14]。其中 ε 为轴向陀螺漂移, ∇_N 为北向加速度计漂移, φ_U 为初始方位误差角, δV_N 为初始速度误差, ω_{ie} 为地球自转周期,周期振荡时间为 24h , ω_s 为舒拉周期,周期为 84.4min , L 为当地纬度, R 为地球半径。

表 1 各误差源对纬度误差的影响情况

Tab. 1 Several error sources influence to latitude error

误差源	传播形式	误差源	传播形式
轴向陀螺漂移	$\left[-\frac{\cos L}{\omega_{ie}} + \frac{\cos L}{\omega_{ie}} \cdot \cos(\omega_{ie} t) \right] \cdot \varepsilon$	北向加速度计漂移	$\frac{\nabla_N}{R(\omega_z^2 - \omega_s^2)} [\cos(\omega_s t) - \cos(\omega_z t)]$
初始方位误差角	$-\varphi_U [\cos L \cdot \sin(\omega_{ie} t) + \frac{\omega_{ie}}{\omega_s} \cos L \cdot \sin(\omega_s t)]$	初始北向速度误差	$\frac{\delta V_N \cdot \sin(\omega_s t)}{R \cos L \cdot g}$

从表 1 中可以看出,初始方位误差角和初始北向速度误差对纬度的影响规律为正弦函数,陀螺和加速度计漂移对纬度的影响规律为余弦函数。经过 12h 导航,对纬度的影响为正弦函数的项变为零,故不用考虑初始方位误差角和初始北向速度误差的影响。而加速度计漂移经过 4h 初始对准后可以精确测定,对纬度误差的影响很小。故可以认为经过 4h 初始对准 12h 导航后产生的误差主要是由于轴向陀螺漂移引起,所以选择 12h 后的纬度误差对轴向陀螺漂移进行辨识。从图 1 中可以看出,系统的纬度误差是一个非线性函数,因此我们考虑采用 RBF 神经网络对轴向陀螺漂移进行辨识。

具体方法为:单轴旋转惯性导航系统经过 4h 的初始对准后,精确地估计出水平陀螺的漂移和加速度计零偏后对其进行修正,然后转入导航状态,经过 12h 导航后,观测系统的纬度变化量,采用 RBF 神经网络进行精确辨识,扣除掉轴向陀螺漂移后再次转入导航状态,进行系统导航解算。

2.2 热态下轴向陀螺常值漂移辨识

系统热态主要是指在对准过程中系统温度变化较小,一般要求温度变化量小于 3℃,系统处于相对稳定的状态。RBF 神经网络系统的输入是惯导系统 12h 输出的纬度误差,RBF 神经网络系统的输出是轴向陀螺漂移。经过单轴旋转后,系统的主要误差源为轴向陀螺漂移,因此轴向陀螺漂移真值的确定方法为:首先不修正轴向陀螺漂移,进行导航实验,实验结束后逐步调整轴向陀螺漂移的计算值,在 72h 内使得定位误差达到最小,以此作为轴向陀螺漂移的真值。为了达到 1nm/72h 的定位精度,需要对轴向陀螺漂移的辨识精度优于 0.0005°/h,为了提高网络训练的精度,将训练误差设置为 0.0003°/h,径向基函数的分布密度 SPREAD 设置为 2,网络训练样本如表 2 所示,当地纬度为 28.222°,训练过程如图 2 所示,当达到一定训练次数后,训练误差变化非常缓慢,趋于收敛。

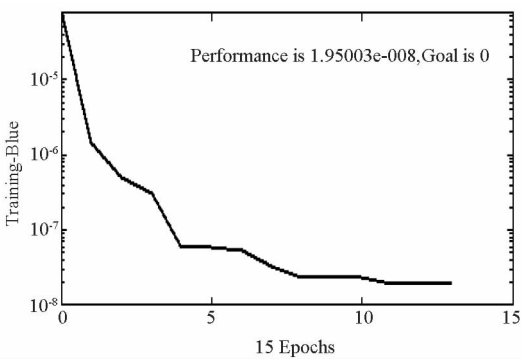


图 2 RBF 神经网络训练过程
Fig. 2 Training curve of the RBF neural network

将一组没有参加网络训练的实验数据作为检验样本,利用训练后的网络进行仿真,计算出网络输出值与实验数据之间的偏差,用以评价网络的逼近精度和泛化能力。表 3 给出了 3 组检验样本经过 RBF 神经网络训练后的结果,图 3 给出了表 3 中第一组辨识前后的定位误差曲线。通过对比可以看出,轴向陀螺漂移辨识误差优于 0.0003°/h,辨识后系统的定位精度大大提高。

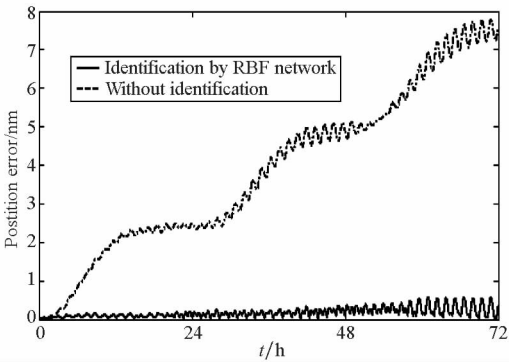


图 3 RBF 神经网络系统辨识前后系统定位误差
Fig. 3 The position error comparison by RBF neural network

2.3 冷态下轴向陀螺常值漂移辨识

系统冷态是指系统启动时完全冷却,启动后温度上升快,温度变化量大,一般大于 10℃。当系统冷态启动进行导航实验时,我们发现,冷态启动相对于热态启动的结果稍差,主要原因是在对准过程中温度发生了较大的变化,轴向陀螺漂移随着温度的变化而改变,在对准结束时利用热态

的 RBF 神经网络系统辨识的轴向陀螺漂移不准确。针对这种情况,设计实验如下:系统冷态启动 16h 对准,要求对准过程中温度变化较大,且温度逐步提升(避免温度出现拐点这种特殊情况,一般情况下,系统温度都是逐步提升的,因此具有一定的普适性),并且在导航过程中温度变化不大,这样可以分离出在对准过程中系统环境温度变化引起的系统导航误差。

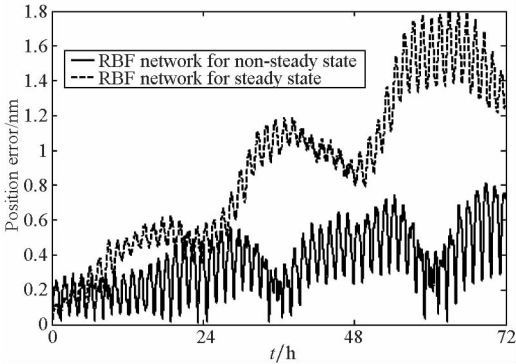


图 4 不同 RBF 神经网络系统辨识后系统定位误差
Fig. 4 The position error by the different RBF neural networks

冷态下轴向陀螺漂移真值的确定方法与热态下轴向陀螺漂移真值的确定方法一致。轴向陀螺漂移经过多温度点补偿后^[15],利用热态 RBF 神经网络进行辨识,结果如表 4 所示,辨识后其导航

定位误差一般为 0.8nm ~ 2.5nm,轴向陀螺漂移误差在 0.0005°/h ~ 0.0015°/h 量级,辨识误差较大。这种轴向陀螺漂移很难通过温度补偿进行消除,因此需要对系统冷态启动时重新建立网络模型,网络输入为系统 12h 输出的纬度误差、温度和温度变化量,输出为轴向陀螺漂移。网络训练样本如表 4 所示,表 5 给出了冷态情况和热态情况下 3 组检验样本经过 RBF 神经网络训练后的结果,图 4 给出了表 5 中第一组实验中系统冷态启动采用不同神经网络辨识的定位误差曲线。可以看出,综合考虑对准时间内温度和温度变化量后,辨识后系统的定位精度大大提高。因此单轴旋转惯导系统中轴向陀螺辨识时要根据系统的状态,选择不同的神经网络系统,对其进行精确辨识。

3 结 论

RBF 人工神经网络能够实现对复杂非线性映射关系的高度逼近,为单轴旋转惯导系统轴向陀螺的辨识提供了一种有效的方法。本文针对惯导系统热态、冷态两种情况,以系统纬度误差、温度及温度变化量为输入,对轴向陀螺漂移进行了精确辨识,RBF 神经网络辨识精度优于 0.0003°/h,极大地提高了系统的导航精度,具有很高的实用价值。

表 2 热态下 RBF 神经网络训练样本
Tab. 2 Sample of the RBF neural network for steady state

纬度(°)	温度范围(℃)	陀螺漂移(°/h)	纬度(°)	温度范围(℃)	陀螺漂移(°/h)
28.2274	37 ~ 39	0.0007	28.2032	60 ~ 61	-0.00253
28.2238	28 ~ 39	0.00022	28.1966	49 ~ 52	-0.00339
28.2162	55	-0.00076	28.1974	50	-0.00324
28.2184	45	-0.00064	28.2150	45 ~ 47	-0.00084
28.2251	35 ~ 36	0.00028	28.1971	42 ~ 43	-0.0033
28.2403	12 ~ 14	0.0024	28.2033	62	-0.0026
28.2474	48 ~ 49	0.0034	28.1982	59	-0.0031
28.2146	55	-0.00096	28.1978	42	-0.00301

表 3 RBF 神经网络系统辨识结果
Tab. 3 Identification result of the RBF neural network

纬度(°)	温度范围(℃)	RBF 辨识的陀螺漂移(°/h)	陀螺漂移真值(°/h)	辨识误差(°/h)	辨识前定位误差(nm/72h)	辨识后定位误差(nm/72h)
28.1959	44 ~ 45	-0.00324	-0.0034	-0.00016	8	0.5
28.2250	43	0.0005	0.0003	-0.0002	1.4	0.6
28.2350	55 ~ 57	0.0020	0.0022	0.0002	4	0.6

表 4 冷态下 RBF 神经网络训练样本

Tab.4 Sample of the RBF neural network for non-steady state

训练 编号	对准时间的 温度变化(℃)	温差 (℃)	校准时间的 温度变化(℃)	温差 (℃)	纬度(°)	陀螺漂移 真值(°/h)	热态 RBF 辨识结果
1	12 ~ 35	23	25 ~ 35	10	28.2413	0.00165	0.0027
2	42 ~ 54	12	51 ~ 54	3	28.1988	-0.00343	-0.0031
3	42 ~ 62	20	53 ~ 62	9	28.2028	-0.00330	-0.0027
4	32 ~ 48	16	44 ~ 48	4	28.2472	0.00325	0.0031
5	29 ~ 38	9	35 ~ 38	3	28.2260	0.00013	0.0007
6	46 ~ 59	13	53 ~ 59	6	28.1980	-0.00345	-0.0031

表 5 不同 RBF 神经网络系统辨识结果对比

Tab.5 Identification result comparison of the different RBF neural networks

检验 编号	对准时间 温变(℃)	温差 (℃)	校准时间 温变(℃)	温差 (℃)	热态 RBF 模型辨识的 系统精度(nm/72h)	冷态 RBF 模型辨识的 系统精度(nm/72h)
1	25 ~ 57	32	40 ~ 57	17	1.8	0.8
2	15 ~ 40	25	30 ~ 40	10	1.7	0.9
3	42 ~ 57	15	52 ~ 57	5	1.3	0.75

参考文献 (References)

[1] Titterton D H, Weston J L. Strapdown inertial navigation technology [M]. 2nd ed., United Kingdom: Institution of Electrical Engineers, 2004.

[2] Lawrence A. Modern inertial technology: navigation, guidance, and control [M]. New York: Springer-Verlag, Inc., 1998.

[3] 许国祯. 惯性技术手册[M]. 北京: 宇航出版社, 1995.
XU Guozhen. Inertial technology handbook [M]. Beijing: Astronautic Publishing House, 1995. (in Chinese)

[4] Levinson E, Giovanni C S. Laser gyro potential for long endurance marine navigation [C]// IEEE Position Location and Navigation Symposium, 1980; 115 - 129.

[5] Levinson E, Majure R. Accuracy enhancement techniques applied to the marine ring laser inertial navigator (MARLIN)[J]. Navigation: Journal of The Institute of Navigation, 1987, 34(1): 64 - 86.

[6] Joint Staff. Master positioning, navigation, and timing plan [R]. Washington, D. C. 20318 - 0600 2003 CJCS, 2003.

[7] Dewall J, May M B, Carvil J, et al. Ship augmented gravity enhancement (SAGE) [C]// IEEE Position Location and Navigation Symposium, 2006; 36 - 43.

[8] 于旭东. 二频机抖激光陀螺单轴旋转惯性导航系统若干关键技术研究[D]. 长沙: 国防科技大学, 2011.
YU Xudong. Research on some key technologies for single-axis rotation inertial navigation system with mechanically dithered ring laser gyroscope [D]. Changsha: National University of Defense Technology, 2011. (in Chinese)

[9] 袁保伦, 饶谷音. 光学陀螺旋转惯导系统原理探讨[J]. 国防科技大学学报, 2006, 28(6): 76 - 80.
YUAN Baolun, RAO Guyin. On the theory of optical gyro rotating inertial navigation system [J]. Journal of National University of Defense Technology, 2006, 28(6): 76 - 80. (in Chinese)

[10] 张伦东, 练军想, 胡小平. 载体角运动对旋转式惯导系统旋转调制效果的影响[J]. 国防科技大学学报, 2011, 33(4): 152 - 156.
ZHANG Lundong, LIAN Junxiang, HU Xiaoping. The effect of vehicle angle motion on rotation modulation technology for rotating INS [J]. Journal of National University of Defense Technology, 2011, 33(4): 152 - 156. (in Chinese)

[11] 吴赛成, 秦石乔, 王省书, 等. 基于姿态解算的 z 向激光陀螺零偏估计方法[J]. 中国激光, 2010, 37(5): 1209 - 1212.
WU Saicheng, QIN Shiqiao, WANG Shengshu, et al. A new method for the estimate of z-direction ring laser gyro drift based on attitude determination [J]. Chinese Journal of Lasers, 2010, 37(5): 1209 - 1212. (in Chinese)

[12] 徐丽娜. 神经网络控制[M]. 北京: 电子工业出版社, 2009.
XU Lina. Neural networks control [M]. Beijing: Publishing House of Electronics Industry, 2009. (in Chinese)

[13] 张泽旭. 神经网络控制与 MATLAB 仿真[M]. 哈尔滨: 哈尔滨工业大学出版社, 2011.
ZHANG Zhexu. Neural networks control and matlab simulation[M]. Harbin: Harbin Institute of Technology Press, 2011. (in Chinese)

[14] 李东明. 捷联式惯导系统初始对准方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2006.
LI Dongmin. Initial alignment method for strapdown inertial navigation system [D]. Harbin: Harbin Engineering University, 2006. (in Chinese)

[15] 张鹏飞, 王宇, 汤建勋, 等. 机抖激光陀螺多温度点实时温度补偿方法的研究[J]. 兵工学报, 2010, 31(5): 562 - 566.
ZHANG Pengfei, WANG Yu, TANG Jianxun, et al. Research on methods for compensating temperature of mechanically dithered RLG [J]. ACTA Armamentar II, 2010, 31(5): 562 - 566. (in Chinese)

想定驱动的仿真资源发现方法*

蔡 楹,赵鑫业,张 鹏,黄柯棣
(国防科技大学 机电工程与自动化学院,湖南 长沙 410073)

摘 要:为解决在云间共享架构的仿真环境下资源的自动发现问题,研究了仿真资源请求的描述方法及相应的资源发现算法,提出了想定驱动的仿真资源发现方法。方法通过扩展的军事想定描述语言表示仿真资源请求,使用双映射三过滤算法实现仿真资源的自动发现。分析表明,仿真想定驱动的资源发现方法可以实现云间共享架构下仿真运行所需的大量资源的自动发现,并对发现结果具有语用校验功能,对于仿真环境的快速部署具有积极意义。

关键词:云间共享;资源发现;想定驱动;军事想定描述语言;双映射三过滤算法;语用层
中图分类号:TP391.9 **文献标志码:**A **文章编号:**1001-2486(2012)03-0053-05

Scenario driven resource discovery method

CAI Ying,ZHAO Xinye,ZHANG Peng,HUANG Kedi

(College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to achieve automatic discovery of simulation resources in a SAC(Sharing Among Clouds) simulation environment, resource request description method and discovery algorithm were studied, and scenario driven resource discovery method was proposed. By using extended MSDL (Military Scenario Definition Language) as resource request scheme and DM&TF (Double Mapping and Triple Filtering) as resource discovery algorithm, resources for cloud simulation can be automatically discovered. Analysis showed scenario driven resource discovery method was able to automatically discover large numbers of resources needed for a simulation in pragmatic level, which is of great help for rapid configuration of simulation environment.

Key words: SAC(Sharing Among Clouds); resource discovery; scenario driven; MSDL(Military Scenario Definition Language); DM&TF (Double Mapping and Triple Filtering) algorithm; pragmatic level

仿真资源的共享和重用是仿真研究的重要支持技术。随着仿真研究对象的复杂化以及跨地域、跨领域的协作不断深入,资源的共享范围和手段也从最初的局域网内部以文件形式的共享,经由以建模仿真资源库(MSRR)为基础的共享,发展到通过网络服务提供仿真服务的共享模式。近年来以网格和云计算为代表的新型网络计算的兴起给传统的共享模式注入了新的理念,同时也对与之匹配的共享支持技术提出了新的要求。在网络化的仿真资源共享方式下,共享环境呈现出资源数量、类型和来源极端丰富的特点,跨机构协作和共享情况增多,资源的异构概率增大。在这种环境下,资源自动发现技术既是重点也是难点。

以UDDI+WSDL组合的资源注册和发现模式在当前的网络服务发现应用领域占主导地位,但由于其主要基于语法进行匹配搜索,所以难以

保证检索效果。而在研究领域,研究机构对于基于语义的资源发现展开了深入研究,也获得了许多成果,包括DAML-S-UDDI^[1], OWLS-UDDI, OWL-MX^[2], WSMO^[3]。然而以上研究关注的均是对匹配方法的研究,所获得的结果也基于同一个假设:即资源请求能够依照资源表示规范进行表述。而这一点在实际情况中由于种种原因,如需要用户熟悉表述规范,并且耗费大量精力,往往是难以符合要求的。

对仿真而言,考虑到一次实验涉及的资源众多,表述资源请求将是十分繁重又难以保证效果的工作,如果无法解决,势必严重影响资源发现的效率和效果。另一方面,仿真对需求的表述有着天然的优势:仿真模型就是在对研究对象抽象理解的基础上构建的,仿真想定包含了对进行仿真实验所需资源的功能需求和使用语境信息,并且

* 收稿日期:2011-07-21
基金项目:国家自然科学基金资助项目(61074108);国家部委基金资助项目(51304102)
作者简介:蔡楹(1984—),男,江西南城人,博士研究生,E-mail:caiyg_cn@126.com
黄柯棣(通信作者),男,教授,博士,博士生导师,E-mail:Huangkedi@nudt.ndu.cn

由于这是仿真必要的准备工作也不会增加仿真人员的负担,如果能够将其应用于资源请求的表述将有利于资源发现。基于以上原因,本文以云间共享架构^[4]下的仿真资源发现为研究对象,主要研究了仿真想定驱动的资源发现方法,提出了基于扩展 MSDL 的资源请求表示规范,并在此基础上提出了双重映射和三重过滤的资源发现算法。

1 云间共享仿真环境中资源发现的问题分析

云间共享是一种集成网格和云计算思想的混合共享模式^[4]。它基于云计算的思想,将资源以服务的形式按需提供给用户,并借鉴网格中虚拟组织的理念实现仿真服务的共享。共享环境中计算相关的服务和应用相关的服务分离,从而仿真服务可以在不同的计算云、存储云之间动态迁移。共享环境中资源通过注册和发布实现资源的共享,虚拟组织代理管理资源的注册和发现,共享架构如图 1 所示。

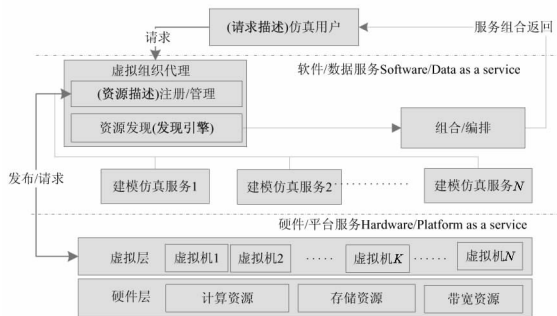


图 1 云间共享仿真环境资源共享架构

Fig. 1 Resource sharing and discovery architecture in SAC based simulation environment

当仿真用户向系统发出资源请求时,代理的资源发现模块将请求信息解析后与已注册的服务信息进行匹配,并将匹配程度最高的资源信息返回给仿真用户。因而资源发现主要研究的是资源的注册信息对资源请求信息的匹配和满足情况,需要解决的问题主要包括 3 个部分:注册资源如何描述、资源请求如何表述、以及资源发现引擎的设计和实现。其中:注册资源的描述由资源提供方负责,需要提供资源的功能、性能、接口、使用前提等有助于资源发现的信息;资源请求由用户进行表述,用于描述用户对资源的需求信息;资源发现引擎由仿真虚拟组织代理负责。三者松散耦合,三方的互操作是进行资源发现的基础。

在协同环境下互操作可分为 3 个层次:语法层、语义层和语用层^[8]。其中语法互操作保证了消息被接受和解析,语义互操作保证了操作双方

对消息理解的一致性,而语用互操作保证了消息接受方能够按照发送方的意图进行操作。三者的互操作能力依次提高,实现难度也逐渐加大。在云间共享架构中计算服务和仿真服务是分离的。计算类型服务的资源发现由于关注点不同(更加侧重于定位、动态特性和服务质量),通常在语法层或者通过简单的语义即可实现,这在网格和云计算领域已有许多成果^[3,6-7]。而仿真服务侧重于发现合适可用的功能,需要在语用层次判别服务的匹配性,两者的研究是正交的,后者是本文研究的重点。

语用发现是在一定语境下进行语义识别和匹配,其中语境的提供是实行语用发现的前提。然而由于语境包含在资源请求中,通常由资源使用方提供,而用户的描述千差万别往往难以合理描述出语境,同时也难以保证资源发现模块能对用户的请求描述做出正确的理解。另一方面,想定是对其包含的元素状态和随后行动规划的特定描述,不仅包含了仿真运行所需模型的需求信息,也对使用语境和条件进行了描述和限制,利用描述规范的想定支持语用层次的仿真资源自动发现具有可行性。

2 基于想定的资源请求规范化描述

在资源网络化共享的建模仿真环境下,想定的开发与资源/模型的开发通常是分离的。如果想定的格式不规范,描述的内容不充分,那么不能保证资源请求信息的完整性,从而也就不能保证资源发现的合理性,因此规范的想定是作为仿真资源请求描述的基础。军事想定描述语言(Military Scenario Definition Language, MSDL)正是基于想定的重用和互操作而提出,由仿真互操作标准组织产品开发组开发,并被标准化的想定规范。

2.1 MSDL 规范分析

MSDL 是一种基于 XML 的军事想定语言规范,目的在于为建模与仿真组织提供验证和载入军事想定的通用机制,并促进联邦仿真的想定一致性^[9]。

MSDL 规范设计为多层的可扩展结构,顶层主要包括 9 大组件,其中:ScenarioID 用于标识想定;Options 用于选择想定中使用的标准和数据模型;Environment 用于描述想定中相关环境的概要信息;ForceSides 用于描述作战派别和军事力量信息;Organizations 用于描述任务或想定相关的组织和装备;Overlays 用于管理控制措施描述的情

报信息;Installations 用于说明任务或想定相关的特定装配信息;TacticalGraphics 用于描述军事想定的控制措施;MOOTWGraphics 用于描述与想定相关的非战争军事行为。

MSDL 组件的语义关系如图 2 所示。资源发现所需的仿真运行相关模型需求信息主要包含在 Organizations 组件中,分为 Units 和 Equipments 两部分。MSDL 通过 Enumerations 将 Unit 或 Equipment 与特定的模型建立关联。Units 定义了想定的作战单元,包括资源的唯一标识(ObjectHandle)、类别信息(SymbolIdentifier)、关系(relation)、资源描述细节(UnitSymbolModifier)以及粒度和聚合情况信息(Model)。Equipment 定义了想定中所有的装备信息,结构和 Units 类似,在此不再赘述;模型的交互定义主要包含在 Overlay 组件中,TacticalGraphics 组件通过子组件 Owner 和 AssociatedOverlays 将作战派别与相应的管理措施层绑定,实际也就是指定单元/装备所属的交互类。此外,MSDL 规范还引用了其他规范作为语义互操作的补充说明,可以充分保证语义上进行互操作^[9]。

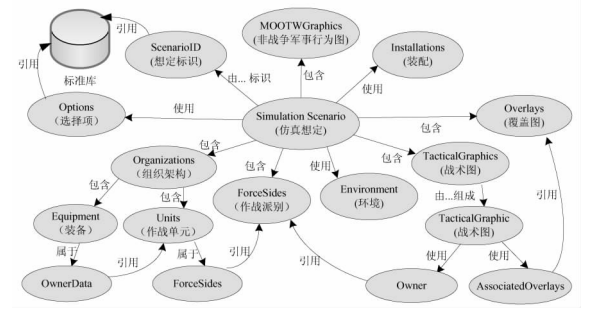


图2 MSDL 组件语义关系

Fig.2 MSDL elements semantic relation

然而,想定信息并不能独立完成从想定输入到仿真运行的转化,需要增加必要的运行配置信息,才能实现整个转化过程的自动进行。因此,为实现云计算仿真环境中从想定输入到仿真执行的自动化,有必要对 MSDL 规范进行扩展。

2.2 基于扩展 MSDL 的资源发现请求规范

传统的仿真过程从想定到仿真执行需要经历 5 个阶段:想定开发(领域专家)、仿真想定生成(仿真专家)、仿真脚本生成、仿真资源配置(仿真实验管理),最后才能运行。从仿真想定到仿真脚本转化的过程就是资源固化的过程,模型被挑选出来并通过索引信息绑定。但实际这是建立在仿真体系架构和运行相关的标准已经确定的基础上的。因此,要实现在云间共享的仿真环境中资源与索引绑定的自动进行,需要增加运行信息的

定义。

资源发现请求基于 MSDL 扩展的规范,如图 3 所示。

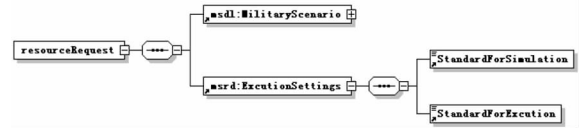


图3 基于 MSDL 扩展的仿真资源请求规范

Fig.3 Extended MSDL for simulation resource request

资源发现请求规范包含 2 个顶层组件:msdl: MilitaryScenario 和 msrd: ExecutionSettings。其中前者为 MSDL 规范,后者为基于 MSRD (Modeling and Simulation Resource Discovery) 仿真运行信息规范。运行信息规范包含 StandardForSimulation (仿真执行标准)和 StandardForExecution (运行相关标准)两个基本组件,并允许扩展其他可选信息,如仿真平台,但扩展部分需引用标准文档并添加必要语义信息。举例来说,进行一次仿真可以选择仿真的标准为 HLA 1516,根据对计算性能的要求申请 4 个节点 CPU,单节点 CPU 为双核 2.8GHz。

仿真标准的确定有利于仿真人员对仿真执行的控制,同时也能够缩小资源搜索范围。运行标准则保证了仿真运算执行的性能,同时也是云计算按需获取资源思想的体现。从而,基于扩展的 MSDL 已经包含了仿真资源发现的必要信息,可以作为资源发现请求规范。

3 想定驱动的资源发现

本文以 OWL-S (Web Ontology Language for Services) 为资源描述规范,以扩展的 MSDL 为资源请求规范,研究想定驱动的资源发现。首先对网络共享仿真环境下资源匹配进行定义。

3.1 仿真中资源发现的匹配定义

由于通过资源发现获得的资源将直接配置用于仿真实验,部分匹配将视为匹配失败,也即认为经典的服务匹配分类^[11]中的“包含相似(subsume)”与“不相似(fail)”对于仿真系统等效。为此,本文将仿真中资源匹配分为 3 类模式,具体定义如下:

- 1) 精确匹配: $output(A) = output(R) \parallel output(R) \text{ subclass of } output(A)$
- 2) 插件式匹配: $output(A) \text{ subsumes } output(R)$
- 3) 不匹配: 不属于以上两种匹配范畴,则失败。

其中 A 代表发布的资源, R 代表服务请求, $output$ 代表资源接口的输出。当发布资源的输出和请求的输出等价时为精确匹配;当请求输出是发布输出的邻接基类 (subclass) 时, 同样判断为精确匹配。当发布输出包含请求输出时, 判断为插件式匹配。当不属于精确匹配和插件式匹配时, 判断为不匹配。匹配程度为: 精确匹配 > 包含性匹配 > 不匹配。

举例来说, 联合作战部队包括陆军、海军、空军和二炮。设 $output(A) = \text{联合作战部队}$:

若 $output(R) = \text{联合作战部队} = output(A)$, 那么判断对资源 A 和请求资源 R 精确匹配; 若 $output(R) = \{\text{陆军、海军、空军和二炮}\}$, 也即 $output(R)$ subclass of $output(A)$, 仍然判断为精确匹配; 若 $output(R) = \text{陆军}$, 那么 $output(A)$ subsumes $output(R)$, 判断为插件式匹配。而如果 $output(R) = \text{联合作战部队}$, 而 $output(A) = \text{陆军}$, 判断为不匹配。

3.2 DM & TF 匹配算法

根据资源发现的需求, 通过双重映射和三重过滤 (Double Mapping & Triple Filtering method, DM & TF) 实现资源发现, 步骤如下:

1) 本体映射。为实现想定开发者与资源发现系统的互操作, 必须确保操作双方在一致理解的基础上进行。将基于扩展 MSDL 规范的想定文件输入系统由想定解析器解析, 通过本体间 (想定开发本体至资源描述本体) 映射, 将想定对资源的需求信息映射成为资源描述本体表述的请求。

2) 语法层过滤。资源请求信息发送至资源注册中心, 对领域、类别、功能等基本信息以及 IO 接口进行语法过滤。对前者的匹配涉及的 MSDL 描述参数包括: Operations 组件中的 Units/Equipment 子组件包含的 (Units/Equipment) SymbolModifiers 类和 Resolution 描述; 对 IO 接口的语法匹配主要对参数个数和数据类型进行匹配, 涉及的 MSRD 描述参数包括 StandardForSimulation 和 StandardForExecution。

3) 语义层过滤。语法层过滤后的结果进行 IO 接口语义匹配和过滤。涉及的参数包括 Unit 单元的 ForceSideHandle 关联的 Overlays 交互信息。

4) 语用层过滤。对语义过滤的结果在语用层次进行验证。主要在战术层检验想定中战术包含的组件能否实现预定功能。涉及的 MSDL 描述参数包括: TacticalGraphics、Overlays 和

Organizations。

5) 对通过检验的结果根据综合相似度从高到底进行排序, 将想定中定义资源的 ScenarioID \rightarrow UUID 属性映射到相似度最高的资源对应的使用关联信息 (grounding)。最后将映射结果发送到共享环境的虚拟组织代理以供仿真实验使用。

3.3 系统设计与应用

构建仿真资源发现系统如图 4 所示。

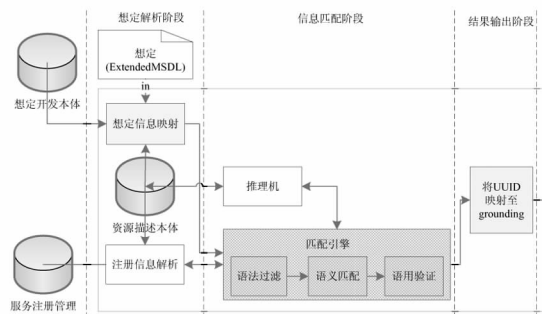


图 4 仿真资源发现架构

Fig. 4 Simulation resource discovery architecture

系统主要包括 3 个部分: 外部输入、解析器和资源匹配引擎。其中, 外部输入包括想定输入以及注册资源信息输入, 解析器包括想定信息解析器、注册信息解析器以及资源描述本体, 资源匹配引擎包括语法过滤器、语义过滤器、语用过滤器和推理机。资源发现通过想定解析、信息匹配和结果输出 3 个阶段完成。

3.4 应用举例

1) 准备

选择 SISO 提供的 SampleMSDL^[10] 为想定样本, 配置项 StandardForSimulation 选择 HLA, 生成仿真资源请求。备选资源基于 OWL-S 描述, 资源描述本体根据 MSDL 本体进行开发和扩展。

2) 步骤

(a) 本体映射。由于资源描述本体与需求表述本体一致, 故本步骤可跳过;

(b) 语法过滤。提取其中的 Unit 和 EquipmentItem 信息用于模型需求信息描述。以 Unit 为例: 提取 Unit 项请求信息, 对 UnitSymbolModifiers 类 (类别信息), Resolution 属性 (粒度信息) 与发布资源信息进行匹配, 过滤语法不匹配资源, 返回发布资源与请求资源语法相似度;

(c) 基于语义的过滤。提取 Unit (EquipmentItem) 的交互信息。以 Unit 为例查找 Unit 所属交互层: 通过 Unit \rightarrow ForceSideHandle 获得所属的作战派别, 在 TacticalGraphics 组件中搜

索 ForceOwnerHandle 属性等于 ForceSideHandle 的选项,并查找该项的 AssociatedOverlays 信息,通过索引到 Overlays 相应的条目即可获得相应的交互信息。提取交互信息的 IO 语义信息与发布资源进行匹配,过滤语义不匹配的资源,返回发布资源与请求资源的语义相似度;

(d)基于语用的过滤。提取 TacticalGraphics 组件的 ForceSideHandle 和 OverLayHandle,遍历 Organizations,找出 ForceSideHandle 包含的 Unit 和 EquipmentItem,基于 TaskSymbolModifiers 类信息通过推理机对结果组合进行校验,过滤其中不匹配的资源组合,返回资源组合的语用相似度。

(e)资源映射。将相似度排名最高的一组结果返回,将 Organizations 组件中的 UUID 属性映射到资源描述对应的 grounding,返回映射结果。

3)结果

实验返回综合匹配最高的一组(760 个)资源集合,对应于想定样本中的 212 个 Unit 和 548 个 EquipmentItem。

4 结 论

本文提出了想定驱动的仿真资源发现方法,通过扩展 MSDL 规范作为资源发现请求规范,基于 DM&TF 算法实现语用层次的资源发现,具有以下优点:

1)仿真想定包含了仿真实验所需的资源(模型等)的需求描述,可以实现一体化的资源搜索,而不需要对每个资源进行单独请求和搜索,返回的资源组合是全局最优的结果;

2)资源匹配基于统一的搜索标准,返回的资源之间能够按想定的交互要求实现互操作;

3)想定信息可以帮助从语用层次上校验搜索结果;

4)整个过程可以自动完成。

综上所述,想定驱动的仿真资源发现方法具有可行性。由于方法的效果与语用层的本体和推理相关,应用于实际系统还需要对发现本体和推理知识库进一步完善。

参考文献(References)

[1] 张童,刘云生,查亚兵. 基于本体的仿真服务定制与组合[J]. 国防科技大学学报,2007,29(4): 105 - 109.
ZHANG Tong, LIU Yunsheng, ZHA Yabing, An ontology based simulation service customization and composition [J]. Journal of National University of Defense Technology, 2007, 29 (4): 105 - 109. (in Chinese)

[2] Matthias K, Benedikt F, Katia S. Automated semantic Web service discovery with OWLS-MX [C]// AAMAS 2006, Hakodate, Japan: 915 - 922.

[3] Keller U, Lara R, Lausen H, et al. Automatic location of services[C]//Proceedings of the 2nd European Semantic Web Conference, LNCS 3532, 2005.

[4] 蔡楹,杨妹,黄柯棣. 基于混合模式的仿真资源共享[J]. 系统仿真学报, 2011, 23:89 - 93.
CAI Ying, YANG Mei, HUANG Kedi. Resource sharing architecture for simulation based on hybrid pattern[J]. Journal of System Simulation, 2011, 23:89 - 93. (in Chinese)

[5] Tamani E, Evripidou P. A pragmatic methodology to Web service discovery [C]// ICWS 2007, Salt Lake City, Utah, USA: IEEE Computer Society, 2007: 1168 - 1171.

[6] Vozmediano R M. A hybrid mechanism for resource/service discovery in Ad-hoc grids [J]. Future Generation Computer Systems, 2009, 25, 717 - 727.

[7] Katagiri T, Kise K, Tangpongprasisit K, et al. A time-to-live based reservation algorithm on fully decentralized resource discovery in grid computing [J]. Parallel Computing, 2005, 31: 529 - 543.

[8] Zeigler B P, Fulton D, Hammonds P. et al. Framework for M&S based system development and testing in net-centric environment[J]. ITEA Journal, 2005, 26(3).

[9] SISO. Standard for: Military scenario definition language (MSDL), SISO-STD-007-2008 [S/OL]. http://www.sisostds.org/digitalLibrary.aspx?command=core_Download&EntryId-29782 last access June 10, 2011.

[10] SISO. SampleMSDL [DB/OL]. http://www.sisostds.org/digitalLibrary.aspx?command=core_Download&EntryId-31559 last access June 10, 2011.

[11] Paolucci M, Kawmura T, Payne T, et al. Semantic matching of web services capabilities[C]// Proceedings of the First International Semantic Web Conference. Sardinia: Springer-Verlag,2002: 333 - 347.

基于设计参数综合评价决策的维修性设计方案优选方法*

罗旭,葛哲学,杨拥民

(国防科学技术大学 装备综合保障技术重点实验室,湖南 长沙 410073)

摘要: 针对维修性设计方案的评估优选问题进行分析,给出一种改进的设计参数综合评价优选方法。

考虑到维修性设计因素评价的不确定性和模糊性的特点,采用集对分析理论对维修性设计因素的模糊语言值评价进行处理,并将层次赋权方法引入方案贴近度评估决策模型,以提高方案优选的可信性。最后以某航天器电源系统的维修性设计方案优选为例说明了该方法的应用及有效性。

关键词: 维修性;设计方案;评价优选;集对分析;贴近度

中图分类号: TP202 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0058-05

Preferred selection method for maintainability design schemes based on integrative evaluation decision-making of design factors

LUO Xu, GE Zhexue, YANG Yongmin

(Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology Changsha 410073, China)

Abstract: Aiming at evaluation and selection of maintainability design schemes, a new preferred selection method based on integrative evaluation of design factors was proposed. Considering the uncertainty and fuzziness of design factors' evaluation, linguistic assessment values of maintainability design factors were quantified on the basis of the set pair analysis theory, and evaluation and decision-making model was established by introducing the weighting method based on AHP into similarity scale of schemes. Finally, preferred selection for design schemes of some manned spacecraft power system was taken as an example to illustrate the application and effectiveness of this method.

Key words: maintainability; design scheme; evaluation and selection; set pair analysis; similarity scale

维修性是由结构设计所赋予的一种质量属性,它是衡量装备维修可行性、有效性和经济性的重要技术指标。由于维修性设计过程中考虑的设计因素较多,而且每种因素具有多种不同的实现途径,导致可能存在若干可行的维修性设计方案。因此,针对已有的多种维修性设计方案,需要综合维修性设计因素进行评估决策,以选择最优的维修性设计方案。

针对维修性设计方案的评价优选问题,Wani和Gandhi分析了维修性各因素之间的相互关系,建立了因素之间的有向网络图及评价矩阵对维修性设计方案进行评估决策^[1];黎放等建立了基于改进灰色分析法的维修性设计评价模型,定量地对维修性设计方案进行评估决策^[2];Corneliu-Alexandru Slavila等应用模糊数学方法,根据维修性定性指标评价得到维修性设计方案的模糊数评价决策值^[3];Chen等应用矢量投影的方法对机械系统设计方案的维修性进行评估决策^[4]。由于

在以上维修性设计方案评价优选方法中,维修性设计因素评价大多是以核对表打分法得到的确定数值形式表现的,忽略了核对评价的不确定性以及人类思维的模糊性,导致决策优选结果的可信度降低,而且由于评估决策模型不能明确维修性设计的薄弱环节,从而影响了维修性设计水平的进一步提高。

因此,本文针对维修性设计因素评价的不确定性和模糊性,研究一种改进的设计参数综合评价决策方法以对维修性设计方案进行优选,同时指出方案设计的相对薄弱环节以激发改进方案的产生。

1 维修性设计因素

影响装备维修性的设计因素较多,根据维修性设计的一般性原则^[5-6],总结归纳出以下7种评价装备维修性的设计因素。

(1) 维修简易性

* 收稿日期:2011-11-18

基金项目:国家自然科学基金项目(51005238)和国家部委预研项目(020201)

作者简介:罗旭(1984—),男,湖北当阳人,博士研究生,E-mail:luoxu2002@sina.com

葛哲学(通信作者),男,副教授,博士,E-mail:gzx@nudt.edu.cn

维修简易性是指装备设计在满足功能要求的前提下,尽可能减少零部件数量,并具有良好工艺性技术支持,以简化维修人员的维修工作,降低对人员的技能要求。

(2) 维修可达性

维修可达性是指接近维修部件进行修理、更换或保养的相对难易程度。要求装备维修部位可看的见、够得着,且具有足够的维修操作空间。

(3) 模块化与互换性

模块化与互换性设计要求将功能相对独立的部分从产品中分离出来,形成能够完全互换的、具有物理连接特性的功能模块,且模块之间具有良好的互换性接口,以便于设备的检测与维修。

(4) 拆卸装配性

拆卸装配性是指通过对产品单元之间的装配连接拆卸方式和序列进行设计,以最大限度地减少拆装时间和难度。对需要更换的设备,还需要有即插即用功能的支持。

(5) 维修安全性

维修安全性是指避免维修人员伤亡或产品损坏的一种设计特性。装备维修性设计应采取措施防止维修事故的发生,保证维修活动的安全。

(6) 人素工程

人素工程要求在维修环境中,维修人员的生理因素、心理因素和人体几何尺寸与装备的关系是合理的、科学的,借以提高维修工作效率和质量、减轻人员疲劳等。

(7) 防差错设计

防差错设计是指在装备结构设计时采取措施(防差错结构、识别标志等)消除维修差错的可能性,减少维修人员维修时的差错。

2 基于设计参数综合的维修性评估方法

2.1 建立设计方案语言值评价决策矩阵

假设有 n 种待选择的维修性设计方案,记为方案集 $X = \{x_1, x_2, \dots, x_n\}$,选取 m 个维修性设计因素指标,记为设计因素集 $Q = \{q_1, q_2, \dots, q_m\}$ 。则方案集中的方案针对设计因素集中各因素的语言值评价构成了优选问题语言值评价决策矩阵 $S = [s_{ij}]_{n \times m}$, $s_{ij} \in L$, s_{ij} 为方案 i 维修性设计因素指标 j 的评价, $L = \{\text{很差, 差, 较差, 一般, 较好, 好, 很好}\}$ 为七标度语言评价集,其中:很差 < 差 < 较差 < 一般 < 较好 < 好 < 很好。

2.2 基于集对分析量化语言值决策矩阵

针对维修性设计因素的模糊等级语言评价,

引入集对分析方法,通过建立实际评价与完美值之间的集对,对设计因素评价进行量化处理。

① 集对分析^[7-8]

集对分析(Set Pair Analysis, SPA)是由我国学者赵克勤提出的一种处理由模糊、信息不完全等原因所致的不确定性问题的系统分析方法,其重视信息处理中的相对性和模糊性,从问题本身分离出相对确定性信息和相对不确定性信息,用联系数进行统一处理。集对分析联系数是集对某一特性上联系的分离定量刻画,定义为

$$\mu = a + b i_0 + c j_0 \tag{1}$$

其中: $a, b, c \in [0, 1]$ 为实数, a 表示同一程度, b 表示差异不确定程度, c 表示对立程度,三者满足 $a + b + c = 1$; i_0 和 j_0 仅为系数标记。

② 语言值矩阵转化为模糊数矩阵^[9-10]

根据模糊集合理论,将设计因素语言值评价转换成 $[0, 1]$ 区间上的三角模糊数,即:“很差” = $(0, 0, 0.16)$, “差” = $(0, 0.16, 0.33)$, “较差” = $(0.16, 0.33, 0.5)$, “一般” = $(0.33, 0.5, 0.66)$, “较好” = $(0.5, 0.66, 0.83)$, “好” = $(0.66, 0.83, 1)$, “很好” = $(0.83, 1, 1)$ 。则可得语言值评价决策矩阵 $S = [s_{ij}]_{n \times m}$ 对应的模糊数决策矩阵为 $\tilde{S} = [\tilde{s}_{ij}]_{n \times m}$, $\tilde{s}_{ij} = (s_{ij}^1, s_{ij}^2, s_{ij}^3)$ 。

③ 模糊数矩阵转化为联系数矩阵

构建评价值 $\tilde{s}_{ij} = (s_{ij}^1, s_{ij}^2, s_{ij}^3)$ 与完美值 $(1, 1, 1) = 1$ 之间的集对 $H = (\tilde{s}_{ij}, 1)$, 将模糊数形式的设计因素指标评价 \tilde{s}_{ij} 转化为联系数形式 μ_{ij} 。

$$\mu_{ij} = a_{ij} + b_{ij} i_0 + c_{ij} j_0 \tag{2}$$

其中: $a_{ij} = s_{ij}^1$, 表示指标值 \tilde{s}_{ij} “能达到完美”的程度; $b_{ij} = s_{ij}^3 - s_{ij}^1$, 表示指标值 \tilde{s}_{ij} “不确定”的程度; $c_{ij} = 1 - s_{ij}^3$, 表示指标值 \tilde{s}_{ij} “不能达到完美”的程度。这样即得联系数形式的方案评价决策矩阵 $\mu = [\mu_{ij}]_{n \times m}$, 以便定量评估决策。

2.3 基于 AHP 法确定维修性设计因素权重

维修性设计因素权重 $W = [w_1, w_2, \dots, w_m]$,

其中 $w_j \geq 0$ 且 $\sum_{j=1}^m w_j = 1$ 。考虑装备维修的特点,根据各设计因素之间的相对重要性,建立设计因素重要性判断矩阵 $A = [a_{ij}]_{m \times m}$, 其中, a_{ij} 表示因素 q_i 相对因素 q_j 的重要性, $a_{ij} = \frac{1}{a_{ji}}$, a_{ij} 采用 1 ~ 9 比较标度方法表示。

为保证设计因素权重确定的合理性,需要对矩阵 A 中各元素 a_{ij} 估计的一致性进行检验。计算矩阵的一致性指标 $CI = \frac{\lambda_{\max} - m}{m - 1}$, 若矩阵的随机

一致性比率 $CR = \frac{CI}{RI} < 0.1$, 认为判断矩阵 A 的估计具有满意的一致性, 否则需要调整判断矩阵 A , 直到取得满意的一致性为止。矩阵随机一致性比率公式中 RI 为同阶矩阵的平均随机指标, 表 1 列出了 1 ~ 9 阶矩阵的 RI 值。

表 1 RI 值

Tab. 1 The value of average random consistency (RI)									
m	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

当判断矩阵 A 的估计具有满意的一致性时, 其最大特征值 λ_{\max} 对应的特征向量 V 经归一化处理得到的向量 W , 即各设计因素的权重向量。

2.4 基于方案贴近度的综合评估决策

根据方案集的评价决策矩阵, 确定正(负)理想决策方案 $x_0^+(x_0^-)$ 作为决策参考。正理想决策方案 x_0^+ 的评估向量为 $\mu^+ = [\mu_1^+, \mu_2^+, \cdots, \mu_m^+]$, 其中 $\mu_j^+ = \max_i \mu_{ij} = a_j^+ + b_j^+ i_0 + c_j^+ j_0$ 。负理想决策方案 x_0^- 的评估向量为 $\mu^- = [\mu_1^-, \mu_2^-, \cdots, \mu_m^-]$, 其中 $\mu_j^- = \min_i \mu_{ij} = a_j^- + b_j^- i_0 + c_j^- j_0$ 。

贴近度是两个群体相似性程度的描述。对两个联系数 $x = a_x + b_x i_0 + c_x j_0, y = a_y + b_y i_0 + c_y j_0$, 可定义它们之间的贴近度:

$$F(x, y) = \frac{a_x a_y + b_x b_y + c_x c_y}{\sqrt{a_x^2 + b_x^2 + c_x^2} \cdot \sqrt{a_y^2 + b_y^2 + c_y^2}}$$

(3)

根据联系数贴近度定义, 计算方案 x_i 与正负理想方案 $x_0^+(x_0^-)$ 在第 j 个设计因素 q_j 下的贴近度 $F(\mu_{ij}, \mu_j^+)$ 和 $F(\mu_{ij}, \mu_j^-)$:

$$\begin{cases} F(\mu_{ij}, \mu_j^+) = \frac{a_{ij} a_j^+ + b_{ij} b_j^+ + c_{ij} c_j^+}{\sqrt{a_{ij}^2 + b_{ij}^2 + c_{ij}^2} \cdot \sqrt{(a_j^+)^2 + (b_j^+)^2 + (c_j^+)^2}} \\ F(\mu_{ij}, \mu_j^-) = \frac{a_{ij} a_j^- + b_{ij} b_j^- + c_{ij} c_j^-}{\sqrt{a_{ij}^2 + b_{ij}^2 + c_{ij}^2} \cdot \sqrt{(a_j^-)^2 + (b_j^-)^2 + (c_j^-)^2}} \end{cases}$$

(4)

考虑各设计因素的权重, 即可得到方案 x_i 与正负理想方案 $x_0^+(x_0^-)$ 之间的方案贴近度:

$$\begin{cases} F(x_i, x_0^+) = \sum_{j=1}^m w_j \cdot F(\mu_{ij}, \mu_j^+) \\ F(x_i, x_0^-) = \sum_{j=1}^m w_j \cdot F(\mu_{ij}, \mu_j^-) \end{cases}$$

(5)

同时, 定义设计因素 q_j 对方案 x_i 综合评价决策的离差影响值:

$$F(\mu_{ij}) = w_j \cdot [F(\mu_{ij}, \mu_j^-) - F(\mu_{ij}, \mu_j^+)]$$

(6)

在 $F(x_i, x_0^+)$ 和 $F(x_i, x_0^-)$ 的基础上, 考虑到决策者的决策态度, 引入决策因子 α , 计算方案决策综合值 $F(x_i)$:

$$F(x_i) = \alpha F(x_i, x_0^+) + (\alpha - 1) F(x_i, x_0^-)$$

(7)

其中: $\alpha \in (0, 1)$, $\alpha > 0.5$ 时, 决策者在决策中持乐观态度; $\alpha = 0.5$ 时, 决策者在决策中持中立态度; $\alpha < 0.5$ 时, 决策者在决策中持悲观态度。

方案决策综合值 $F(x_i)$ 越大, 说明该方案越靠近正理想方案且远离负理想方案, 其维修性越优, 则 $F(x_i)$ 最大的方案即为维修性最优方案。

设计因素对方案决策的离差影响值 $F(\mu_{ij})$ 越大, 说明该设计因素对方案决策的负面影响越大, 则 $F(\mu_{ij})$ 最大者对应的设计因素 q_j 即为方案 x_i 维修性设计的相对薄弱环节。

3 案例应用分析

以某航天器电源系统为分析对象, 对已有的 4 种维修性设计方案进行评估决策, 选取维修性最优方案。考虑维修简易性、维修可达性、模块化与互换性、拆卸装配性、维修安全性、人素工程和防差错设计 7 个方面的设计因素, 根据专家分析得到 4 个方案在各设计因素下的模糊语言评价, 如表 2 所示, 即构成语言值评价决策矩阵。

表 2 方案集维修性设计因素的语言值评价

Tab. 2 Linguistic assessment values of maintainability design factors							
	维修 简易性	维修 可达性	模块化与 互换性	拆卸 装配性	维修 安全性	人素 工程	防差错 设计
方案一	较差	一般	好	较好	好	较好	较好
方案二	较好	一般	一般	一般	好	一般	一般
方案三	较好	较好	较好	较好	较好	较好	好
方案四	一般	好	好	好	较好	一般	较差

3.1 获取联系数形式评价决策矩阵

用 2.2 节的量化方法, 将表 2 对应的语言值评价决策矩阵量化为联系数形式的评价决策矩阵 μ , 见式(8)。

$$\mu = \begin{bmatrix} 0.16 + 0.34i + 0.5j & 0.33 + 0.33i + 0.34j & 0.66 + 0.34i & 0.5 + 0.33i + 0.17j & 0.66 + 0.34i & 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j \\ 0.5 + 0.33i + 0.17j & 0.33 + 0.33i + 0.34j & 0.33 + 0.33i + 0.34j & 0.33 + 0.33i + 0.34j & 0.66 + 0.34i & 0.33 + 0.33i + 0.34j & 0.33 + 0.33i + 0.34j \\ 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j & 0.5 + 0.33i + 0.17j & 0.66 + 0.34i \\ 0.33 + 0.33i + 0.34j & 0.66 + 0.34i & 0.66 + 0.34i & 0.66 + 0.34i & 0.5 + 0.33i + 0.17j & 0.33 + 0.33i + 0.34j & 0.16 + 0.34i + 0.5j \end{bmatrix}$$

(8)

3.2 设计因素赋权

根据7个方面的设计因素对航天器电源系统维修性水平的影响程度进行分析比较,构建设计因素重要性判断矩阵A:

$$A = \begin{bmatrix} 1 & 1/2 & 2 & 2 & 1/3 & 1/2 & 3 \\ 2 & 1 & 3 & 3 & 1/2 & 1 & 4 \\ 1/2 & 1/3 & 1 & 1 & 1/5 & 1/3 & 2 \\ 1/2 & 1/3 & 1 & 1 & 1/5 & 1/3 & 2 \\ 3 & 2 & 5 & 5 & 1 & 2 & 6 \\ 2 & 1 & 3 & 3 & 1/2 & 1 & 4 \\ 1/3 & 1/4 & 1/2 & 1/2 & 1/6 & 1/4 & 1 \end{bmatrix}$$

计算重要性判断矩阵A的随机一致性比率CR,对A中各元素的估计进行一致性检验。由 $\lambda_{\max} = 7.0554$, $CI = \frac{\lambda_{\max} - m}{m - 1} = \frac{7.0554 - 7}{7 - 1} = 0.009$, $RI = 1.32$,得 $CR = \frac{CI}{RI} = 0.007 < 0.1$ 。

说明重要性判别矩阵A具有满意的一致性。求解判断矩阵最大特征值 $\lambda_{\max} = 7.0554$ 对应的特征向量:

$$V = [0.251 \ 0.418 \ 0.146 \ 0.146 \ 0.733 \ 0.418 \ 0.093]^T$$

归一化求解,得各设计因素权重向量:

$$W = [0.114 \ 0.190 \ 0.066 \ 0.066 \ 0.332 \ 0.190 \ 0.042]^T$$

3.3 确定理想方案评估向量

根据方案集的联系数形式评价矩阵,分别确定正理想决策方案 x_0^+ 和负理想决策方案 x_0^- 的评估向量 μ^+ 和 μ^- 。

$$\mu^+ = [\mu_1^+, \mu_2^+, \dots, \mu_7^+] = [0.5 + 0.33i + 0.17j$$
$$0.66 + 0.34i \ 0.66 + 0.34i \ 0.66 + 0.34i$$
$$0.66 + 0.34i \ 0.5 + 0.33i + 0.17j \ 0.66 +$$
$$0.34i]$$
$$\mu^- = [\mu_1^-, \mu_2^-, \dots, \mu_7^-] = [0.16 + 0.34i + 0.5j$$
$$0.33 + 0.33i + 0.34j \ 0.33 + 0.33i + 0.34j$$
$$0.33 + 0.33i + 0.34j \ 0.5 + 0.33i + 0.17j$$
$$0.33 + 0.33i + 0.34j \ 0.16 + 0.34i + 0.5j]$$

3.4 计算方案贴近度决策优选

根据式(4)和(5),计算该航天器电源系统维修性设计方案 x_i 与正(负)理想决策方案 x_0^+ (x_0^-)之间的综合贴近度,如表3所示。

表3 设计方案 x_i 与 x_0^+ (x_0^-)的综合贴近度
Tab.3 Integrative similarity scale between x_i and x_0^+ (x_0^-)

	x_1	x_2	x_3	x_4
$F(x_i, x_0^+)$	0.9187	0.8725	0.9715	0.9400
$F(x_i, x_0^-)$	0.9384	0.9496	0.9055	0.9176

取决策因子 $\alpha = 0.5$,由式(7),计算各个方案的决策综合值 $F(x_i)$,如下:

$$F(x_1) = -0.009848, F(x_2) = -0.038574$$
$$F(x_3) = 0.033013, F(x_4) = 0.011175$$

可见, $F(x_3) > F(x_4) > F(x_1) > F(x_2)$,该航天器电源系统维修性设计方案的优劣排序依次是方案三,方案四,方案一,方案二,故选择方案三为最优维修性设计方案。

通过专家对各个设计方案设计因素的语言值评价,可以发现,相对于方案一、方案二和方案四,方案三每个设计因素评价均较好,特别是对电源系统总体维修性有较大影响的设计因素,即权重较大的维修性设计因素,其评价基本上是4个备选方案中最好的,因此选择方案三作为电源系统最优的维修性设计方案是合理的。同时,选择方案三也符合试验研究的情况,充分说明了本文研究的维修性设计方案优选方法的有效性。

对于方案三,根据式(6)计算设计因素对方案决策的离差影响值 $F(\mu_{3j})$,可得:

$$\max_{1 \leq j \leq m} F(\mu_{3j}) = F(\mu_{35}) = 0.0145$$

可见设计因素 q_5 (维修安全性)是航天器电源系统方案三维修性设计的相对薄弱环节。通过维修安全性设计改进,可使方案三的总体维修性水平进一步有效提高。

4 结 论

针对维修性设计方案优选问题,提出一种改进的设计参数综合评价决策方法。该方法采用集对分析理论将维修性设计因素定性评价转化为定量描述,同时考虑设计因素的权重,使得方案贴近度评估决策模型对方案的优劣排序更加科学合理,从而为维修性设计方案优选提供一种有效的方法。

参考文献 (References)

[1] Wani M F, Gandhi O P. Development of maintainability index for mechanical system[J]. Reliability Engineering and System Safety, 1999, 65: 259 – 270.

[2] 黎放, 杨元, 杨建军. 装备维修性设计参数综合评估方法[J]. 工程设计学报. 2008, 15(1): 11 – 16.

LI Fang, YANG Yuan, YANG Jianjun. Integrative evaluation on parameters of maintainability design of mechanical equipment[J]. Journal of Engineering Design, 2008, 15(1): 11 – 16. (in Chinese)

[3] Slavila C A, Decreuse C, Ferney M. Fuzzy approach for maintainability evaluation in the design process [J]. Concurrent Engineering, 2005, 13: 291 – 299.

[4] Chen L, Cai J G. Using vector projection method to evaluate maintainability of mechanical system in design review [J]. Reliability Engineering and System Safety, 2003, 81: 147 – 154.

[5] 国防科学技术工业委员会. GJB/Z91—1997 维修性设计手册[S]. 北京: 中国标准出版社, 1997.

The committee of National Defense Scientific and Technical Industry. GJB/Z91 – 1997 maintainability design manual[S]. Beijing: Standards Press of China, 1997. (in Chinese)

[6] 甘茂治, 吴真真. 维修性设计与验证[M]. 北京: 国防工业出版社, 1995.

GAN Maozhi, WU Zhenzhen. Maintainability design an evaluation [M]. Beijing: Defense Industry Press, 1995. (in Chinese)

[7] 赵克勤. 集对分析及其初步应用[M]. 杭州: 浙江科学技术出版社, 2000.

ZHAO Keqin. Set pair analysis and its preliminary applications [M]. Hangzhou: Zhejiang Science and Technology Press, 2000. (in Chinese)

[8] 杨俊杰, 周建中, 方仍存, 等. 基于集对分析的不确定多属性决策方法[J]. 控制与决策. 2008, 23(12): 1423 – 1426.

YANG Junjie, ZHOU Jianzhong, FANG Rengcun, et al. Uncertain multi-attribute decision making methods based on set pair analysis [J]. Control and Decision, 2008, 23(12): 1423 – 1426. (in Chinese)

[9] Chiou H K., Tzeng G H. Fuzzy multiple-criteria decision-making approach for industrial green engineering [J]. Environmental Management. 2002, 30(6): 816 – 830.

[10] 刘英平, 杨素君, 陈晓霞. 不确定信息的机械产品设计方案评价方法[J]. 机械科学与技术. 2008, 27(7): 946 – 949.

LIU Yingping, YANG Sujun, CHEN Xiaoxia. Study of evaluation method for mechanical product design schemes with uncertainty information [J]. Mechanical Science and Technology for Aerospace Engineering. 2008, 27(7): 946 – 949. (in Chinese)

(上接第 43 页)

参考文献 (References)

[1] Mohammad M, Fangming D, Naiffer R, et al. Increased flexural modulus and strength in SWNT/epoxy composites by a new fabrication method [J]. Polymer, 2006, 47 (1): 293 – 298.

[2] Varley R J, Hodgkin J H, Hawthorne D G. Toughening of a trifunctional epoxy system; Part VI [J]. Polymer, 2001, 42: 3847 – 3858.

[3] 李晓丹, 王维, 郭淑齐, 等. 耐高温树脂的固化动力学分析及其力学性能[J]. 纤维复合材料, 2009, 24(4): 24 – 26.

LI Xiaodan, WANG Wei, GUO Shuqi, et al. The analysis of curing kinetics of heat resistant resin and research of its composites performances [J]. Fiber Composites, 2009, 24(4): 24 – 26. (in Chinese)

[4] 代晓青, 肖加余. 等温 DSC 法研究 RFI 用环氧树脂固化动力学研究 [J]. 复合材料学报, 2008, 25 (4): 18 – 23. (in Chinese)

DAI Xiaqing. XIAO Jiayu. Curing kinetics of epoxy resin for RFI process using isothermal DSC [J]. Acta Materiate Compositae Sinica, 2008, 25(4): 18 – 23. (in Chinese)

[5] Barton J M. The application of differential scanning calorimetry (DSC) to the study of epoxy resin curing reactions [J]. Advances in Polymer Science, 1985, 72: 1 – 11.

[6] 谢旻, 顾轶卓, 李敏, 等. 碳纤维/双马树脂预浸料固化过

程动态力学性能[J]. 复合材料学报, 2010, 27(4): 52 – 58.

XIE Min, GU Yizhuo, LI Min, et al. Dynamic mechanical analysis of the curing of carbon fiber/bismaleimide resin prepreg [J]. Acta Materiate Compositae Sinica, 2010, 27(4): 52 – 58. (in Chinese)

[7] Yun N G, Won Y W, Kim S C. Toughening of carbon fiber/epoxy composite by inserting polysulfone film to form morphology spectrum [J]. Polymer, 2004, 45(20): 6953 – 6958.

[8] Mimura K, Ito H, Fujioka H, Improvement of thermal and mechanical properties by control of morphologies in PES modified epoxy resins [J]. Polymer, 2000, 41 (12): 4451 – 4459.

[9] Catalani A, Bonicelli M G. Kinetics of the curing reaction of a diglycidyl ether of bisphenol A with a modified polyamine [J]. Thermochimica Acta, 2005, 438(1): 126 – 129.

[10] Li P, Yang X P, Yu Y H, et al. Cure kinetics, microheterogeneity, and mechanical properties of the high-temperature cure of vinyl ester resins [J]. Journal of Applied Polymer Science, 2003, 92(2): 1124 – 1133.

[11] Kamal M R, Sourour S. Kinetics and thermal characterization of thermoset cure [J]. Polymeric Materials Science and Engineering, 1973, 13: 59 – 64.

基于影响网络与不完全信息多阶段博弈的
作战行动序列模型及求解方法^{*}

杜正军, 陈超, 姜鑫

(国防科技大学 信息系统工程重点实验室, 湖南 长沙 410073)

摘要: 拟制作战计划的关键是生成作战行动序列。考虑作战行动过程中的不确定性以及激烈对抗性, 以影响网络分析作战行动和战场态势之间的相互影响, 用不完全信息博弈分析敌对双方之间的对抗, 建立了基于影响网络和不完全信息多阶段博弈的作战行动序列模型, 并给出了求解方法。通过计算实例说明了模型及其求解方法。结果显示该模型生成的 COA 更适应于对抗环境。

关键词: 作战行动序列; 影响网络; 不完全信息多阶段博弈

中图分类号: **文献标志码:**A **文章编号:**1001 - 2486(2012)03 - 0063 - 05

Modeling and solution method of course of action based on
influence net and multi-stage games with incomplete information

DU Zhengjun, CHEN Chao, JIANG Xin

(Science and Technology on Information Systems Engineering Laboratory,
National University of Defense Technology, Changsha 410073, China)

Abstract: The development of Course of Action (COA) is a key step of military planning. In most existing literature on COA development, the previous studies only take unilateral actions of friendly force into account. Considering the antagonism of war, we propose models that contain actions of both friendly and enemy force based on multi-stage games with incomplete information. Influence net is used to analyze the uncertain relationships between actions and battlefield situation. In the end, numerical examples are presented to illustrate the models and solution, showing that the COA developed in the current study is suitable for the antagonistic situation.

Key words: course of action; influence net; multi-stage games with incomplete information

作战计划拟制的实质是对军事资源进行分配, 合理规划作战行动序列 (Course of Action, COA), 以改变战场态势达到目标状态。Haider 等^[1-6]用影响网络分析行动与态势之间的相互影响并优化行动序列。Wagenhals 等^[7]基于影响网络和 Petri 网开发了 CAESAR II/EB 用于 COA 的评估。Rafi 等^[8]提出了 ATIN (Activation Timed Influence Net) 模型用于对 COA 建模与优化。Zhang 等^[9]采用有限自动机和形式化语言来描述行动方案的逻辑特征以及定量特征, 并开发了一种形式化的建模工具 COAST 用于 COA 的调度。Pousi^[10]运用多准则影响图对 EBO 计划过程进行分析。Yaman 等^[11]运用模糊认知图支持 EBO 的计划过程。Wilkins^[12]、Myers^[13]、Tate^[14-15]分别建立了基于 HTN 的军事计划系统用于辅助 COA 的生成。目前对 COA 的研究主要是应用人工智能和自动规划领域的技术解决作战行动规划问

题, 在计划过程中考虑己方作战行动对战场态势的影响, 通过己方一系列的作战行动实现战场态势的转变, 未考虑到敌方可能采取针对性的作战行动对战场态势的影响。

战争关系着参战双方的生死存亡, 双方都期望战场态势按照己方的意愿发展, 而战场态势的演变取决于双方作战行动的共同作用。在制定作战计划时应该考虑到这种对抗, 这就需要在生成作战行动序列的过程中考虑敌我双方的博弈。其次由于参战双方会极力隐藏己方的真实信息以及战争本身的不确定性, 一方对另一方的兵力、武器装备以及作战目标等缺乏完全信息, 从博弈论的角度讲, 就是对另一方的可行策略、支付函数等博弈参数不确定, 即军事对抗活动是不完全信息博弈。通常情况下, 作战目标通过多个阶段的作战行动完成, 上一阶段博弈的结果是下一阶段博弈的条件, 双方博弈的目标是使最终的战场态势更

^{*} 收稿日期: 2011 - 11 - 28
基金项目: 国家自然科学基金资助项目 (71101149); 国防科技大学优秀研究生创新资助项目 (B100502)
作者简介: 杜正军 (1982—), 男, 四川南充人, 博士研究生; E-mail: outstandingdzj@163. com;
陈超 (通信作者), 男, 讲师, 博士, E-mail: chenchao@gmail. com

有利于己方。基于此,本文从作战双方博弈的角度,用不完全信息多阶段博弈分析作战行动序列的生成。影响网络便于描述行动与效果之间的关系,提供了一种有效描述和分析作战行动序列的框架。在目前应用影响网络对作战行动序列的研究工作中^[1-8],并未考虑到资源约束以及双方博弈。本文在影响网络的基础上考虑资源约束以及敌我双方的博弈,建立了一个基于影响网络和不完全信息多阶段博弈的作战行动序列模型。

1 模型及求解方法

1.1 模型描述

文献[2-6,8]利用影响网络模型分析了作战行动与战场态势之间的相互影响,模型中的节点分为状态节点和作战行动节点,但只考虑了单方的作战行动对战场态势的影响。本模型中考虑红蓝双方的作战行动节点,战场态势由红蓝双方的作战行动共同决定。将作战行动过程视为不完全信息多阶段博弈,共有 K 个阶段,在博弈的每一阶段,红蓝双方根据资源约束确定己方和敌方的可行行动,并通过观察对方所选择的行动,根据贝叶斯法则来修正对敌方类型的概率分布,然后选择己方的行动。下面从状态向量、作战行动、作战行动与状态之间的相互影响、资源约束、类型空间以及行动与行动空间 6 个方面对模型进行描述。

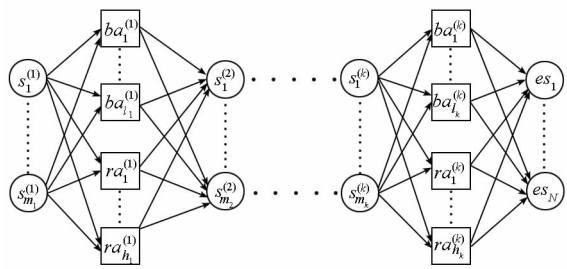


图 1 考虑双方博弈的影响网络 COA 模型

Fig. 1 COA models based on influence net considering actions of both sides

(1) 状态向量

第 k ($1 \leq k \leq K$) 阶段的状态向量 $\mathbf{s}^{(k)} = (s_1^{(k)}, \dots, s_{m_k}^{(k)})$, 最终阶段的状态向量 $\mathbf{s}^e = (s_1^e, \dots, s_{N_e}^e)$ 。第 k 阶段的状态向量既是第 k 阶段行动的前提条件,也是第 $k-1$ 阶段行动的效果。状态向量的每一维分量为 $\{0, 1\}$ 变量, 0 表示该情况不出现, 1 表示该情况出现。

(2) 作战行动项

作战行动项为最基本的任务单元,在作战过程中的每一阶段,双方都有一些作战行动项可供

选择。设第 k 阶段红方共有 h_k 项可选作战行动 $\{a_{r,1}^{(k)}, \dots, a_{r,h_k}^{(k)}\}$, 蓝方共有 l_k 项可选作战行动 $\{a_{b,1}^{(k)}, \dots, a_{b,l_k}^{(k)}\}$ 。

(3) 作战行动与状态之间的相互影响

第 k 阶段状态向量为 $\mathbf{s}^{(k)}$ 的情况下,红方第 i 项作战行动成功的概率为 $p(a_{r,i}^{(k)} | \mathbf{s}^{(k)})$, 蓝方第 j 项作战行动成功的概率为 $p(a_{b,j}^{(k)} | \mathbf{s}^{(k)})$; 红方的行动为 $\varphi_i^{(k)}$, 蓝方的行动为 $\psi_j^{(k)}$ 的条件下,第 $k+1$ 阶段第 i 维状态向量 $s_i^{(k+1)}$ 的概率为 $p(s_i^{(k+1)} | \varphi_i^{(k)}, \psi_j^{(k)})$ 。

①第 k 阶段蓝方第 i 项作战行动的成功概率为:

$$p(a_{b,i}^{(k)} | \mathbf{s}^{(k)}) = \begin{cases} p(a_{b,i}^{(k)}) + hb_i(s^{(k)})(1-p(a_{b,i}^{(k)})) & \text{if } hb_i(s^{(k)}) \in [0, 1] \\ p(a_{b,i}^{(k)}) + hb_i(s^{(k)})p(a_{b,i}^{(k)}) & \text{if } hb_i(s^{(k)}) \in [-1, 0] \end{cases}$$

②记第 k 阶段蓝方的作战行动 $(a_{b,1}^{(k)}, \dots, a_{b,l_k}^{(k)})$ 以及红方的作战行动 $(a_{r,1}^{(k)}, \dots, a_{r,h_k}^{(k)})$ 为 $\mathbf{a}^{(k)}$, $\mathbf{a}^{(k)}$ 对状态向量 $\mathbf{s}^{(k)}$ 的第 i 维分量 $s_i^{(k)}$ 的影响参数为 $gs_i^{(k)}(\mathbf{a}^{(k)})$, $s_i^{(k)}$ 的概率为:

$$p(s_i^{(k)} | \mathbf{a}^{(k)}) = \begin{cases} p(s_i^{(k)}) + gs_i^{(k)}(\mathbf{a}^{(k)})(1-p(s_i^{(k)})) & \text{if } gs_i^{(k)}(\mathbf{a}^{(k)}) \in [0, 1] \\ p(s_i^{(k)}) + gs_i^{(k)}(\mathbf{a}^{(k)})p(s_i^{(k)}) & \text{if } gs_i^{(k)}(\mathbf{a}^{(k)}) \in [-1, 0] \end{cases}$$

(4) 资源约束

在实际的作战过程中,指挥员能调用的资源是有限的。设第 k 阶段蓝方的资源为 $\{s_{b,1}^{(k)}, \dots, s_{b,c_k}^{(k)}\}$, 红方的资源为 $\{s_{r,1}^{(k)}, \dots, s_{r,d_k}^{(k)}\}$ 。第 k 阶段蓝方第 i 项作战行动对第 j 种资源的需求为 $bn_{ij}^{(k)}$, 红方第 i 项作战行动对第 j 种资源的需求为 $rn_{ij}^{(k)}$ 。

(5) 类型空间

红蓝双方的类型空间分别为 $\Theta_r = \{\theta_{r,i}\}$, $\Theta_b = \{\theta_{b,j}\}$, $i \in \mathbf{N}, j \in \mathbf{N}$ 。给定类型空间 $(\theta_{r,i}, \theta_{b,j})$ 联合概率分布 $P(\theta_{r,i}, \theta_{b,j})$ 的情况下,红方根据己方的类型可以推测出蓝方的类型的概率分布 $P(\theta_{b,j} | \theta_r)$, θ_r 为红方实际的类型。同理,蓝方也可以根据 θ_b 推测出红方的类型分布 $P(\theta_{r,i} | \theta_b)$ 。第 k 阶段红方对蓝方所属类型的信念为 $P_r^{(k)} = (p_{r,1}^{(k)}, \dots, p_{r,i}^{(k)}, \dots, p_{r,N_b}^{(k)})$, 蓝方对红方所属类型的信念为 $P_b^{(k)} = (p_{b,1}^{(k)}, \dots, p_{b,j}^{(k)}, \dots, p_{b,N_r}^{(k)})$ 。

(6) 行动、行动空间

①第 k 阶段红方的行动指在 h_k 项作战行动项中选择其中满足资源约束的 n 项 ($0 \leq n \leq h_k$) 执行,定义为向量 $(a_{r,1}^{(k)}, \dots, a_{r,h_k}^{(k)})$, 向量中的分量为 $\{0, 1\}$ 变量, 0 表示不执行该作战行动项, 1 表示执行该作战行动项,记为 $\varphi_i^{(k)}$ 。红方所有的可行行动构成了第 k 阶段红方的行动集 $\Phi^{(k)} = \{\varphi_1^{(k)}, \varphi_2^{(k)}, \dots, \varphi_{f_k}^{(k)}\}$, $f_k \leq 2^{h_k}$ 。同理定义蓝方第 k

阶段的行动 $\psi_j^{(k)}$ 以及行动集 $\Psi^{(k)} = \{\psi_1^{(k)}, \psi_2^{(k)}, \dots, \psi_{g_k}^{(k)}\}$, $g_k \leq 2^{l_k}$ 。

②红方第1阶段行动集 $\Phi^{(1)}$ 到第 K 阶段的行动集 $\Phi^{(K)}$ 的笛卡尔积 $\Phi^{(1)} \times \Phi^{(2)} \times \dots \times \Phi^{(K)}$ 构成了红方的行动空间 Φ , 即 $\Phi = \Phi^{(1)} \times \Phi^{(2)} \times \dots \times \Phi^{(K)}$ 。同理定义蓝方的行动空间 Ψ , $\Psi = \Psi^{(1)} \times \Psi^{(2)} \times \dots \times \Psi^{(K)}$ 。

1.2 数学模型

综上,作战行动序列可以用六元组 $G = \langle W, \Omega, U, \Theta, P, H \rangle$ 表示,其中:

(1) $W = \{R, B\}$ 表示局中人,即红蓝双方;

(2) $\Omega = \{\Phi, \Psi\}$ 表示红蓝双方的行动空间;

① $\Phi = \Phi^{(1)} \times \Phi^{(2)} \times \dots \times \Phi^{(K)}$

$\Phi^{(k)}$ 的任意一项可行行动 $\varphi_m^{(k)}$ 须满足资源约束:

$$\sum_{i=1}^{h_k} r a_i^{(k)} \times r n_{ij}^{(k)} \leq r s_j^{(k)}$$

$$m = 1, \dots, f_k; j = 1, \dots, c_k; k = 1, \dots, K$$

② $\Psi = \Psi^{(1)} \times \Psi^{(2)} \times \dots \times \Psi^{(K)}$

$\Psi^{(k)}$ 的任意一项可行行动 $\psi_n^{(k)}$ 须满足资源约束:

$$\sum_{i=1}^{l_k} b a_i^{(k)} \times b n_{ij}^{(k)} \leq b s_j^{(k)}$$

$$n = 1, \dots, g_k; j = 1, \dots, d_k; k = 1, \dots, K$$

(3) $H = \{H_r^{(k)}, H_b^{(k)}\}$, 表示第 k 阶段的行动历史,即在作战行动过程中双方实际执行的行动。 $H_r^{(k)} = \{\varphi^{(1)}, \varphi^{(2)}, \dots, \varphi^{(k-1)}\}$, $H_b^{(k)} = \{\psi^{(1)}, \psi^{(2)}, \dots, \psi^{(k-1)}\}$, $1 < k \leq K$

(4) $\Theta = \{\Theta_r, \Theta_b\}$ 表示类型空间,即红蓝双方可能的类型的集合,红方共有 N_r 种类型,蓝方共有 N_b 种类型; $\Theta_r = \{\theta_{r,1}, \theta_{r,2}, \dots, \theta_{r,N_r}\}$, $\Theta_b = \{\theta_{b,1}, \theta_{b,2}, \dots, \theta_{b,N_b}\}$

(5) $P = \{P_r^{(k)}, P_b^{(k)}\}$, $P_r^{(k)}$ 表示第 k 阶段红方对蓝方所属类型的信念, $P_b^{(k)}$ 表示第 k 阶段蓝方对红方所属类型的信念,信念需根据作战行动过程中获取的信息进行更新;

① 当 $k = 1$ 时: $P_r^{(1)} = (p_{r,1}^{(1)}, \dots, p_{r,i}^{(1)}, \dots, p_{r,N_b}^{(1)})$, $P_b^{(1)} = (p_{b,1}^{(1)}, \dots, p_{b,j}^{(1)}, \dots, p_{b,N_r}^{(1)})$, $p_{r,i}^{(1)} = P(\theta_{b,i} | \theta_r)$, $p_{b,j}^{(1)} = P(\theta_{r,j} | \theta_b)$, $i = 1, \dots, N_b$, $j = 1, \dots, N_r$

② 当 $1 < k \leq K$: $P_r^{(k)} = (p_{r,1}^{(k)}, \dots, p_{r,i}^{(k)}, \dots, p_{r,N_b}^{(k)})$, $P_b^{(k)} = (p_{b,1}^{(k)}, \dots, p_{b,j}^{(k)}, \dots, p_{b,N_r}^{(k)})$, $p_{r,i}^{(k)} = P(\theta_{b,i} | \theta_r, H_b^{(k)})$, $p_{b,j}^{(k)} = P(\theta_{r,j} | \theta_b, H_r^{(k)})$, $i = 1, \dots, N_b$, $j = 1, \dots, N_r$

(6) U 表示支付函数,由双方类型组合以及策

略确定; $U = f(\alpha_{i_1}^{(k)}, \beta_{j_1}^{(k)}, \theta_{r,i_2}, \theta_{b,j_2})$, $k = 1, \dots, K$, $i_1 = 1, \dots, (\Delta r_k)^{N_r}$, $j_1 = 1, \dots, (\Delta b_k)^{N_b}$, $i_2 = 1, \dots, N_r$, $j_2 = 1, \dots, N_b$

1.3 模型求解

不完全信息动态博弈的解为完美贝叶斯均衡,完美贝叶斯均衡要求在博弈的每一阶段,局中人的均衡策略在每一个信息集开始的后续博弈上构成贝叶斯均衡,并用贝叶斯法则修正关于其他局中人类型的信息^[17]。在本文中,为便于模型求解,将博弈用策略式表示。在博弈的第一阶段,根据类型的先验概率以及赢得矩阵,用海萨尼转换为标准形式,确定双方的均衡策略以及本阶段的行动。在从第二阶段开始的后续博弈中,双方根据对方的行动修正对方类型的后验概率,并确定该阶段的行动。具体求解步骤如下:

(1) 根据资源约束分析双方行动空间,将博弈用策略式表示,针对双方的每一种类型组合,用影响网络计算出赢得矩阵。

(2) 第1阶段,根据先验概率分布 P_{ij} , 以及赢得矩阵 $A_i^{(ij)}$, 用海萨尼转换转化为标准形式。博弈的标准形式 G_1^* 中红方标准化策略是类型空间 Θ_r 到策略的函数,红方的标准化策略 $x_i^{(1)}$ 表示为 $x_i^{(1)} = (\alpha_{i_1}^{(1)}, \alpha_{i_2}^{(1)}, \dots, \alpha_{i_{N_r}}^{(1)})$, 即红方属于第一种类型则采用策略 $\alpha_{i_1}^{(1)}$, 属于第二种类型则采用策略 $\alpha_{i_2}^{(1)}$, 属于第 N_r 种类型则采用策略 $\alpha_{i_{N_r}}^{(1)}$, 同理蓝方的标准化策略表示为 $y_j^{(1)} = (\beta_{j_1}^{(1)}, \beta_{j_2}^{(1)}, \dots, \beta_{j_{N_b}}^{(1)})$ 。标准化赢得矩阵记为 $\Gamma^{(1)}$, $\Gamma^{(1)}$ 的元素 $\Gamma_{ij}^{(1)}$ 的值等于 $\sum_{n=1}^{N_r} \sum_{m=1}^{N_b} p_{mn} \times A_{i_n, j_m}^{(nm)}$, i_n, j_m 分别为标准化策略 $x_i^{(1)}, y_j^{(1)}$ 中对应于红方第 n 种类型的策略的下标以及蓝方第 m 种类型的策略的下标。求解得到红蓝双方的均衡策略,双方根据均衡策略执行第一阶段的行动 $\varphi^{(1)}, \psi^{(1)}$ 。

(3) 第 k ($1 < k \leq K$) 阶段,根据 $k-1$ 阶段的行动历史,用贝叶斯法则修正对方类型的后验概率。蓝方根据红方 $k-1$ 阶段的行动 $\varphi^{(k-1)}$, 修正红方属于类型 $\theta_{r,n}$ ($n = 1, 2, \dots, N_r$) 的后验概率为

$$P(r\theta_n | \varphi^{(k-1)}) = \frac{P(\varphi^{(k-1)} | r\theta_n) \times b p_n^{(k-1)}}{\sum_{i=1}^{N_r} b p_i^{(k-1)} \times P(\varphi^{(k-1)} | r\theta_i)}$$

方根据蓝方 $k-1$ 阶段的行动 $\psi^{(k-1)}$, 修正蓝方属于类型 $\theta_{b,m}$ ($m = 1, 2, \dots, N_b$) 的后验概率为

$$P(b\theta_m | \psi^{(k-1)}) = \frac{P(\psi^{(k-1)} | b\theta_m) \times r p_m^{(k-1)}}{\sum_{i=1}^{N_b} r p_i^{(k-1)} \times P(\psi^{(k-1)} | b\theta_i)}$$

据双方类型的后验概率分布 $P_{ij}^{(k)}$ 以及对应的赢得矩阵 $A_k^{(ij)}$, 用海萨尼转换重新化为标准形式 G_k^* , 红方的标准化策略 $x_i^{(k)}$ 表示为 $x_i^{(k)} = (\alpha_{i_1}^{(k)}, \alpha_{i_2}^{(k)}, \dots, \alpha_{i_{N_r}}^{(k)})$ 。蓝方的标准化策略表示为 $y_j^{(k)} = (\beta_{j_1}^{(k)}, \beta_{j_2}^{(k)}, \dots, \beta_{j_{N_b}}^{(k)})$ 。标准化赢得矩阵记为 $\Gamma^{(k)}$, $\Gamma^{(k)}$ 的元素的 $\Gamma_{ij}^{(k)}$ 值等于 $\sum_{n=1}^{N_r} \sum_{m=1}^{N_b} p_{nm}^{(k)} \times A_{i_n, j_m}^{(nm)}$ 。求解得到红蓝双方的标准化均衡策略。红蓝双方根据均衡策略执行第 k 阶段的行动 $\varphi^{(k)}, \psi^{(k)}$ 。

2 计算实例

2.1 算例

考虑一个三阶段的作战行动序列的网络, 如图 2 所示。

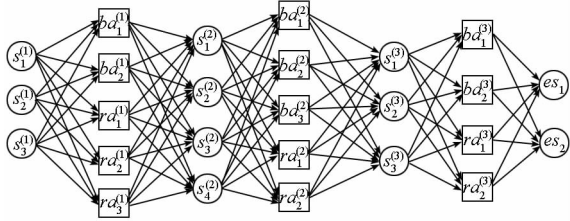


图 2 示例行动影响图

Fig. 2 Example of COA models based on influence net

(1) 每一阶段的状态向量分别为 $(s_1^{(1)}, s_2^{(1)}, s_3^{(1)})$, $(s_1^{(2)}, s_2^{(2)}, s_3^{(2)})$, $(s_1^{(3)}, s_2^{(3)}, s_3^{(3)})$, 最后一阶段状态向量为 (s_1^e, s_2^e) 。

(2) 红方每一阶段可选作战行动项分别为 $\{a_{r,1}^{(1)}, a_{r,2}^{(1)}, a_{r,3}^{(1)}\}$, $\{a_{r,1}^{(2)}, a_{r,2}^{(2)}\}$, $\{a_{r,1}^{(3)}, a_{r,2}^{(3)}\}$, 蓝方每一阶段可选作战行动项为 $\{a_{b,1}^{(1)}, a_{b,2}^{(1)}\}$, $\{a_{b,1}^{(2)}, a_{b,2}^{(2)}\}$, $\{a_{b,1}^{(3)}, a_{b,2}^{(3)}\}$ 。

(3) 随机生成节点之间的影响参数。

(4) 红方每一阶段的资源分别为 $(s_{r,1}^{(1)}, s_{r,2}^{(1)}, s_{r,3}^{(1)})$, $(s_{r,1}^{(2)}, s_{r,2}^{(2)}, s_{r,3}^{(2)})$, $(s_{r,1}^{(3)}, s_{r,2}^{(3)}, s_{r,3}^{(3)})$, 蓝方每一阶段资源分别为 $(s_{b,1}^{(1)}, s_{b,2}^{(1)})$, $(s_{b,1}^{(2)}, s_{b,2}^{(2)}, s_{b,3}^{(2)})$, $(s_{b,1}^{(3)}, s_{b,2}^{(3)}, s_{b,3}^{(3)})$ 。随机生成红蓝双方的资源约束以及作战行动项对资源的需求。

(5) 类型空间及其信念概率。

红蓝双方都有两种类型, 分别记为 $\theta_{r,1}, \theta_{r,2}$ 和 $\theta_{b,1}, \theta_{b,2}$ 。类型的联合概率分布为 $P(\theta_{r,1}, \theta_{b,1}) = 0.1, P(\theta_{r,1}, \theta_{b,2}) = 0.4, P(\theta_{r,2}, \theta_{b,1}) = 0.3, P(\theta_{r,2}, \theta_{b,2}) = 0.2$ 。红方实际类型为 $\theta_{r,1}$, 蓝方实际类型为 $\theta_{b,2}$ 。根据边缘概率公式, 红方对蓝方类型的判断为 $P(\theta_{b,1} | \theta_{r,1}) = 0.2, P(\theta_{b,2} | \theta_{r,1}) = 0.8$ 。蓝方对红方的类型判断为 $P(\theta_{r,1} | \theta_{b,2}) = 2/3, P(\theta_{r,1} | \theta_{b,2}) = 1/3$ 。

2.2 求解

(1) 分析红蓝双方每一阶段的可行行动集

易知蓝方第一阶段满足资源约束的行动集为 $\{(0,1), (1,0), (1,1)\}$; 第二阶段的行动集为 $\{(0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1)\}$; 第三阶段行动集为 $\{(0,1), (1,0)\}$ 。红方第一阶段行动集为 $\{(0,0,1), (0,1,0), (1,0,0), (1,0,1)\}$; 第二阶段行动集为 $\{(0,1), (1,0)\}$; 第三阶段行动集为 $\{(0,1), (1,0), (1,1)\}$ 。

(2) 红蓝双方博弈过程

① 第一阶段

红蓝双方根据先验概率分布 $P = \begin{bmatrix} 0.1 & 0.4 \\ 0.3 & 0.2 \end{bmatrix}$, 以及在各种类型组合下的赢得矩阵 $A_1^{(11)}, A_1^{(12)}, A_1^{(21)}, A_1^{(22)}$, 用海萨尼转换得到标准形式为 $\Gamma^{(1)}$, $\Gamma^{(1)}$ 为一个 $24^2 \times 30^2$ 的矩阵, $\Gamma_{ij}^{(1)} = \sum_{n=1}^2 \sum_{m=1}^2 p_{nm} \times A_{i_n, j_m}^{(nm)}$ 。求解得到

(I) 红方的均衡策略为 $(x_1^{(1)}, x_2^{(1)}) = (0.763, 0.237)$, 其中 $x_1^{(1)} = (\alpha_{11}^{(1)}, \alpha_{12}^{(1)})$, $\alpha_{11}^{(1)} = (0, 1, 0, 0, 1, 1, 1)$, $\alpha_{12}^{(1)} = (1, 0, 0, 0, 1, 0, 1)$, $x_2^{(1)} = (\alpha_{21}^{(1)}, \alpha_{22}^{(1)})$, $\alpha_{21}^{(1)} = (0, 1, 0, 0, 1, 0, 1, 1)$, $\alpha_{22}^{(1)} = (1, 0, 0, 0, 1, 0, 1)$

(II) 蓝方的均衡策略为 $(y_1^{(1)}, y_2^{(1)}) = (0.176, 0.824)$, 其中 $y_1^{(1)} = (\beta_{11}^{(1)}, \beta_{12}^{(1)})$, $\beta_{11}^{(1)} = (1, 0, 0, 1, 1, 0, 1)$, $\beta_{12}^{(1)} = (1, 0, 0, 1, 1, 0, 1)$, $y_2^{(1)} = (\beta_{21}^{(1)}, \beta_{22}^{(1)})$, $\beta_{21}^{(1)} = (1, 0, 0, 0, 1, 0, 1)$, $\beta_{22}^{(1)} = (1, 0, 0, 1, 1, 0, 1)$

分析可知: 红方类型为 $\theta_{r,1}$, 则第一阶段采取行动 $\varphi_1^{(1)} = (0, 1, 0)$, 类型为 $\theta_{r,2}$, 则第一阶段采取行动 $\varphi_2^{(1)} = (1, 0, 0)$ 。红方实际类型为 $\theta_{r,1}$, 因此采取行动 $\varphi^{(1)} = (0, 1, 0)$ 。第一阶段作战行动完成后, 蓝方能判断出红方的类型。 $P(\theta_{r,1} | \theta_{b,2}, \varphi^{(1)}) = 1, P(\theta_{r,1} | \theta_{b,2}, \varphi^{(1)}) = 0$ 。蓝方类型为 $\theta_{b,1}, \theta_{b,2}$ 在第一阶段都采取行动 $\psi^{(1)} = (1, 0)$ 。第一阶段作战行动完成后, 红方从蓝方的行动中未获取到信息更新对蓝方类型的信念。 $P(\theta_{b,1} | \theta_{r,1}, \psi^{(1)}) = 0.2, P(\theta_{b,2} | \theta_{r,1}, \psi^{(1)}) = 0.8$ 。

② 第二阶段

第一阶段行动结束后战场态势 $(s_1^{(2)}, s_2^{(2)}, s_3^{(2)}, s_4^{(2)}) = (0.5651, 0.4449, 0.0372, 0.9914)$ 为第二阶段博弈开始的条件。

(I) 从红方的角度来说, 类型组合为 $(\theta_{r,1}, \theta_{b,1})$ 和 $(\theta_{r,1}, \theta_{b,2})$ 的概率分别是 0.2、0.8。且红方知道蓝方能确知红方的类型为 $\theta_{r,1}$ 。通过影响网络

计算出两种类型组合下的赢得矩阵分别为

$$A_2^{(11)} = \begin{bmatrix} 0.2146 & 0.2952 & 0.2437 & 0.4085 & 0.2119 & 0.2855 & 0.247 & 0.4263 & 0.2202 & 0.3323 \\ 0.1812 & 0.7956 & 0.4043 & 0.9708 & 0.1614 & 0.7806 & 0.4323 & 0.9977 & 0.2327 & 0.8511 \\ 0.5654 & 0.9331 & 0.7732 & 0.9912 & 0.5475 & 0.9278 & 0.8047 & 0.9966 & 0.63 & 0.9408 \\ 0.2045 & 0.2545 & 0.2439 & 0.4098 & 0.2136 & 0.2953 & 0.247 & 0.4265 & 0.211 & 0.2919 \\ 0.1035 & 0.7328 & 0.4063 & 0.9727 & 0.1761 & 0.7953 & 0.4326 & 0.998 & 0.1596 & 0.7892 \\ 0.4911 & 0.9136 & 0.7754 & 0.9916 & 0.5648 & 0.9304 & 0.8051 & 0.9966 & 0.5569 & 0.9235 \end{bmatrix}$$
$$A_2^{(12)} = \begin{bmatrix} 0.814 & 0.8531 & 0.8402 & 0.8768 & 0.8111 & 0.851 & 0.8392 & 0.8805 & 0.8064 & 0.8607 \\ 0.1067 & 0.9471 & 0.265 & 0.9929 & 0.0899 & 0.9429 & 0.261 & 0.9977 & 0.0688 & 0.9546 \\ 0.1287 & 0.9207 & 0.1897 & 0.9918 & 0.123 & 0.913 & 0.1947 & 0.9877 & 0.1345 & 0.8968 \\ 0.8064 & 0.8446 & 0.8402 & 0.877 & 0.81 & 0.8531 & 0.8391 & 0.8805 & 0.8019 & 0.8523 \\ 0.06 & 0.9315 & 0.265 & 0.9933 & 0.0848 & 0.9453 & 0.2606 & 0.9977 & 0.0395 & 0.9405 \\ 0.1084 & 0.9006 & 0.1901 & 0.9916 & 0.1253 & 0.9092 & 0.1947 & 0.9874 & 0.117 & 0.8857 \end{bmatrix}$$

利用海萨尼转换转化为标准形式 $\Gamma^{(2)}$ 。由于红方知道蓝方能确知红方的类型为 $\theta_{r,1}$,而红方不能确知蓝方的类型。故求得红方的策略为 $x^{(2)} = (0, 1, 0, 1)$ 。蓝方的标准化策略为 $y^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)})$, $\beta_1^{(2)} = (0, 1, 1, 0, 1)$, $\beta_2^{(2)} = (1, 0, 1, 0, 1)$

(Ⅱ)蓝方已经确知红方类型为 $\theta_{r,1}$,即已经知道类型组合为 $(\theta_{r,1}, \theta_{b,2})$,则赢得矩阵为 $A_2^{(12)}$ 。对于蓝方,知道红方分析的结果会采取行动 $\alpha^{(2)} = (0, 1, 0, 1)$,而蓝方的均衡策略为 $\beta_2^{(2)} = (1, 0, 1, 0, 1)$ 。

分析可知:红方第二阶段采取行动 $\varphi^{(2)} = (0, 1)$,蓝方第二阶段采取行动 $\psi^{(2)} = (1, 0, 1)$ 。且红方从蓝方的行动中更新对蓝方类型的信念, $P(\theta_{b,1}|\theta_{r,1}, \psi^{(2)}) = 0, P(\theta_{b,2}|\theta_{r,1}, \psi^{(2)}) = 1$ 。

③第三阶段

红蓝双方都知道类型组合为 $(\theta_{r,1}, \theta_{b,2})$,通过影响网络计算出赢得矩阵为

$$A_3 = \begin{bmatrix} 0.8063 & 0.8607 \\ 0.0687 & 0.9546 \\ 0.1344 & 0.8967 \end{bmatrix}$$

求解得出红方会采取行动 $\varphi^{(3)} = (0, 1)$,蓝方行动为 $\psi^{(3)} = (0, 1)$,赢得值为 0.8063。

3 结 论

战争是激烈的对抗活动,战场态势的演变取决于作战双方作战行动的共同作用,于是双方决策、行动的正确性相互依赖。考虑作战计划的对抗性以及不确定性,建立了一个基于多阶段不完全信息博弈的作战行动过程模型,用影响网络分析作战行动与态势之间的相互影响,最后给出了一个计算实例并进行了求解。与只考虑单方作战

行动与战场态势之间影响的方法相比,本文提出的方法更适应于对抗环境,为指挥员制定作战计划提供了一种辅助决策的方法。

参考文献(References)

[1] Rosen J A, Smith W L. Influence net modeling with causal strengths: an evolutionary approach [C]// Command and Control Research and Technology Symposium, Naval Post Graduate School, Monterey, 1996:25 - 28.

[2] Haider S, Zaidi A K, Levis A H. A heuristic approach for best set of actions determination in influence nets [C]// Proceedings of IEEE International Conference on Information Reuse and Integration, Las Vegas ,2004.

[3] Haider S, Zaidi A K. Transforming timed influence nets into time sliced bayesian networks[C]// Proceedings of Command and Control Research and Technology Symposium ,2004.

[4] Haider S, Levis H. Modeling time-varying uncertain situations using dynamic influence net [J]. International Journal of Approximate Reasoning,2008,49(2) :488 - 502.

[5] Haider S, Levis A H. Finding effective courses of action using particle swarm optimization [C]// Proceedings of World Congress on Computational Intelligence, Hong Kong, 2008, 1135 - 1140.

[6] Haider S, Levis A H. Effective courses of action determination to achieve desired effects[J]. IEEE Transactions on System, Man and Cybernetics,2007,37(6) :1140 - 1150.

[7] Wagenhals L W, Levis A H. Modeling support of effects based operations in war games [C]// 7th Command and Control Research and Development Symposium, Naval Post Graduate School, Monterey, Ca, 2002.

[8] Rafi F M, Zaidi A K, Kazakos P P, et al. Optimization of actions in activation timed influence nets [J]. Informatica , 2009, 285 - 296.

[9] Zhang L, Mitchell B, Janczura C. COAST: an operational planning tool for course of action development and analysis [C]// Symposium for the 9th International Command and Control Research and Technology Symposium, 2003.

参考文献 (References)

[1] Goddard J S. Pose and motion estimation from vision using dual quaternion-based extended kalman filtering [D]. Knoxville: the University of Tennessee, 1997.

[2] 丁尚文, 王惠南, 刘海颖, 等. 基于对偶四元数的航天器交会对接位姿视觉测量算法 [J]. 宇航学报, 2009, 30 (6): 2145 – 2150.

DING Shangwen, WANG Huinan, LIU Haiying, et al . Algorithm of vision measure for relative position and pose of RVD spacecrafts based on dual-quaternion [J]. Journal of Astronautics, 2009, 30 (6): 2145 – 2150. (in Chinese)

[3] 张世杰. 基于单目视觉的航天器相对导航理论与算法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2005.

ZHANG Shijie. Study on theory and algorithms of monocular vision-based navigation for spacecraft [D]. Harbin: Harbin Institute of Technology. (in Chinese)

[4] 武元新. 对偶四元数导航算法与非线性高斯滤波研究 [D]. 长沙: 国防科技大学, 2005.

WU Yuanxin. Research on dual-quaternion navigation algorithm and monocular gaussian filtering [D]. Changsha: National University of Defense Technology, 2005. (in Chinese)

[5] Olson C F, Matthies L H, Schoppers M, et al. Rover navigation using stereo ego-motion [J]. Robotics and Autonomous Systems, 2003, 43 (4): 215 – 229.

[6] Allen D W, Johnson N, Alison A P. Vision-aided inertial navigation for flight control [J]. Journal of Aerospace Computing, Information and Communication, 2005, (2): 348 – 360.

[7] Andreas H. Relative position sensing by fusing monocular vision and inertial rate sensors [D]. Stanford: Stanford University, 2003.

[8] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation [C]// Proceedings of the IEEE International Conference on Robotics and Automation, Roma, Italy, 2007.

[9] Stein G P, Mano O, Shashua A. A robust method for computing vehicle ego-motion [C]// Proceedings of the IEEE Intelligent Vehicles Symposium, Dearborn, USA, 2000.

[10] 刘威, 魏存伟, 赵逢, 等. 基于单目视觉的自行车运动参数鲁棒估计 [J]. 机器人, 2009, 31 (1): 20 – 26.

LIU Wei, WEI Cunwei, ZHAO Feng, et al. A robust method for vehicle ego-motion estimation based on monocular vision [J]. Robot, 2009, 31 (1): 20 – 26. (in Chinese)

[11] 李万里. 车载组合导航自适应滤波及抗野值算法研究 [D]. 长沙: 国防科技大学, 2008.

LI Wanli. Research on adaptive Kalman filter and fault-tolerant algorithm used in in-vehicle integrated navigation system [D]. Changsha: National University of Defense Technology, 2008. (in Chinese)

[12] Bäumker M, Heimes F J. New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of an hybrid inertial navigation system [M]. Frankfurt am Main: Bundesamt für Kartographie und Geodäsie, 2002: 197 – 212.

[13] Feng G H, Wu W Q, Cao J L, et al. Algorithm for monocular visual odometry/SINS integrated navigation [J]. Journal of Chinese Inertial Technology, 2011, 19 (3): 210 – 215.

[10] Pousi J. Decision analytical approach to effect based operations [D]. Espoo: Helsinki University of Technology, 2009.

[11] Yaman D, Polat S. A fuzzy cognitive map approach for effect based operations: an illustrative case [J]. Information Sciences, 2009: 382 – 403.

[12] Wilkins D E, Desimone R V. Applying an AI planner to military operations planning [R]. SRI International Research Report, 1993.

[13] Myers K, Tyson W M. PASSAT: A user-centric planning framework [C]// Proceedings of the 3rd International NASA Workshop on Planning and Scheduling for Space. Houston, USA. 2002.

[14] Tate A, Levine J, Jarvis P, et al. Using AI planning technology for army small unit operations [C]// Proceedings of the Fifth International Conference on Artificial Intelligence Planning Systems, 2000.

[15] Potter S, Tate A, Wickler G. Using I-X process panels as intelligent to do lists for agent coordination in emergency response [C]// Proceedings of the Information Systems for Crisis Response and Management, Newark, New Jersey, USA, 2006.

[16] Fudenberg D, Tirole J. Game theory [M]. Massachusetts: The MIT Press, 1991.

[17] John H. Games with incomplete information played by “ bayesian ” players [J]. Management Science, 1967: 159 – 182.

(上接第 67 页)

考虑需求相关的可修复系统备件配置模型^{*}

周伟,蒋平,刘亚杰,郭波

(国防科技大学 信息系统与管理学院, 湖南 长沙 410073)

摘要: 对于一个拥有多台相同设备的可修复系统来说,一种备件的需求以及其对应的备件保障方案将影响着系统的可工作设备数量,从而影响着设备其他备件的需求,故假设备件需求相互独立且与系统自身备件保障方案无关而开展备件配置研究是不合理的。在考虑备件需求相关以及备件的保障方案对备件需求的影响的前提下,利用马尔可夫理论描述了备件需求规律,针对不串件拼修和串件拼修两种情况分别给出了可用度的计算方法,并在此基础上以备件配置费用作为优化目标,以可用度作为约束构建备件优化配置模型,给出了边际分析求解模型的方法。通过案例分析,验证了模型的正确性。

关键词: 备件;需求相关;模型;可用度;拼修策略

中图分类号: TP301 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0068-06

A spares allocation model for repairable systems considering correlated spare demands

ZHOU Wei, JIANG Ping, LIU Yajie, GUO Bo

(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: For a repairable system consisting of multiple equipments of the same type, spare demand and its supply plan will affect the number of workable equipments, and hence exerts effects on the demand of other spares for the equipment. Therefore, the assumption that the spare demands are mutually independent and have no relation with the spare supply plan of the system is not correct. This study considers the correlated demands as well as the effects of the spare supply plan on the spare demand, and describes the spare demand law by the Markov process. Availability calculus methods are presented according to the two cases of cannibalization and non-cannibalization respectively. Based on the methods, a spare optimal allocation model is constructed by setting spare allocation cost as the optimization target and system availability as the constraints. The model is solved by marginal analysis. Finally, a case study is presented to illustrate the validity of the proposed model.

Key words: spare; correlated demands; model; availability; cannibalization policy

随着科技的进步,各种系统越来越复杂,如何保持系统稳定运行是一项重要工作,一旦系统部件失效造成停机,将会面临巨大的损失。维修保障部门必须储备相应的备件,使得系统在发生故障后能够尽快修复,然而备件储备的同时也会带来相应的库存费用、存储失效损失以及其他方面的费用,所以需要备件库存量进行优化,权衡缺货风险与备件配置费用。当前国内外对备件库存优化方面开展了大量的研究,取得了较大进展。一方面,研究者们不再将备件优化局限于单个备件配置等级或者备件配置层级,在多层级多等级备件优化配置方面产生了以 METRIC 模型为代表的一系列研究成果,并得到了较好的应用^[1-6];另一方面,部分研究者开始关注备件配置与维修策略等相关因素的协同优化^[7-9]。但是这些研究都

是基于各类备件需求相互独立的基本前提,它们假定备件的需求并不取决于当前工作设备的数量,并没有意识到系统缺少备件未能及时修复造成系统停机时会影响该系统中的其他部件对备件的需求。当系统停机等待备件进行修复期间,余下的部件停止工作直到相应的备件到来,那就暗示着它们并没有备件需求直到系统修复。基于备件需求相互独立假设的备件需求预测往往会造成误差,不能准确地描述实际的备件需求。例如,因为未考虑备件需求相关的影响,文献[10]中系统的备件短缺量被高估了 37.3%。考虑需求相关的可修复系统备件配置属于时变需求库存控制的一种,当前研究时变需求库存控制文献尽管考虑了需求变化的特点,但是他们的研究对象如生产库存量等并不会影响需求的到达^[11-15],而对于考

^{*} 收稿日期:2011-11-11

基金项目:国家自然科学基金资助项目(70971132)

作者简介:周伟(1985—),男,湖南新邵人,博士研究生, E-mail:zhouwei_nudt@163.com;

郭波(通信作者),男,教授,博士,博士生导师, E-mail:boguo@nudt.edu.cn

考虑需求相关的可修复系统来说,它的备件库存量会通过影响工作设备数量反作用于备件需求。本文研究可修复系统备件配置时考虑了备件需求相关的影响,利用连续马尔可夫链的相关理论分析了可修复系统各类备件需求之间的关系,并以系统稳态可用度为约束,以备件配置费用为目标构建了备件优化配置模型。

1 模型建立

1.1 问题基本假设

现实生活中最典型的可修复系统由数台相同的设备以及其所属的维修保障机构共同构成,这些设备的备件统一由该维修保障机构的备件仓库保障^[16]。本文将针对此类典型的可修复系统,考虑其各种备件需求之间的关系来构建备件优化配置模型,现对模型提出如下假设:

(1)某使用单位有 K 台相同设备和一个备件仓库,每台设备包含 L 种部件,第 l 种部件的机用数为 n_l ,各部件均为关键部件,任意部件失效均会造成系统停机。

(2)部件在工作时间失效率相互独立,不考虑故障从属、故障级联的情况,部件 l 在工作期间的失效率为常数 λ_l ,部件在非工作时间不失效。

(3)部件失效后采用 $(s-1, s)$ 库存策略送修,在有备件的情况下换件维修时间忽略不计,部件 l 的修复率为常数 μ_l 。系统维修能力无限,部件均能修复,且修复如新。

(4)可修复系统为一个封闭系统,不考虑上级仓库备件支援的情况。

1.2 备件需求相关分析

可修件一般采用 $(s-1, s)$ 库存策略,即当设备上某个部件出现故障后,立即将故障件送修,如果该部件存有备件,立即予以更换,如果没有备件,则设备因备件供应延误导致停机,等待有故障件修复返回后设备重新开始运行。在这种策略下,存在一个经典的库存数量守恒函数^[1],即

$$s_l = OH_l + DI_l + BO_l \tag{1}$$

式中, s_l 为部件 l 的备件配置量, OH_l 为部件 l 的现有库存量, DI_l 为部件 l 的在修数, BO_l 为部件 l 的备件短缺量。可修复系统除了存在库存数量守恒函数,同时也存在部件守恒函数,因为对于一个封闭的库存系统来说,同种部件的总量是有限的,它们以可工作设备部件、停机设备上的可工作部件、可用储备部件(可用备件)以及在修件的形态存在,部件守恒函数表述为

$$S_l = WI_l + SI_l + DI_l + OH_l \tag{2}$$

其中 S_l 为系统中部件 l 的数量, WI_l 为处于工作设备上部件 l 的数量, SI_l 为停机待修设备上部件 l 的数量。由于部件只有在工作时才可能失效,所以备件需求到达决定于 WI_l 的数量。 WI_l 的数量取决于当前可工作设备数量 WE , 因为有 $WI_l = n_l \times WE$, 任意一种部件短缺造成设备停机, 均会影响工作设备数, 从而影响系统中所有部件的备件需求, 这就是备件需求相关的体现。

因为备件需求相关的存在, 各时段可工作设备将会发生变化, 系统备件需求到达并不是一个齐次泊松过程。可修复系统不可能任意时刻均处于完好状态, 对于任意备件 $l, l = 1, 2, \dots, L$, 其平均备件需求率必然要小于 $Kn_l\lambda_l$, 但是随着各备件库存数量的增加, 工作设备数量增加, 各类备件需求将增加, 当各备件的配置数量趋向于无穷时, 任意时刻期望工作设备数趋向于 K , 备件需求率趋向于 $Kn_l\lambda_l$, 此时备件需求服从强度为 $Kn_l\lambda_l$ 的齐次泊松过程, 可修复系统备件需求率变化如图 1 所示。基于备件需求相互独立的备件需求预测方法割裂了各类备件需求之间的相互关系, 并不能准确地描述备件需求, 该方法主要是通过帕尔姆定理来计算部件在修数^[1], 但是帕尔姆定理的基本前提是备件需求服从一个稳定的齐次泊松过程, 而这种基本假设由于备件需求相关的存在而并不成立。

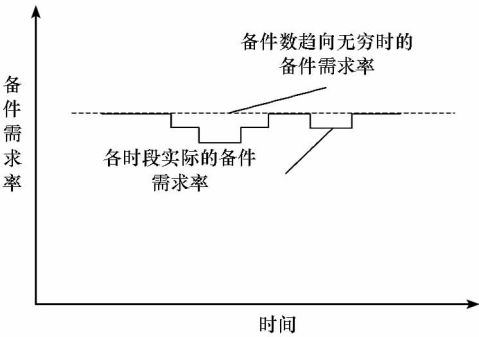


图 1 备件需求率变化
Fig. 1 Spare demand change

1.3 系统稳态可用度

当前国内外主要选用备件满足率或者备件短缺量作为部件级的备件保障效能参数, 而系统级主要选用可用度作为备件保障效能参数^[17]。可用度主要从保持系统稳定运行力度的角度评价备件保障效能。对于单台设备来说, 可用度体现为任意时刻设备未因任何备件短缺而停机的概率。对于一个具备多台相同设备的系统来说, 可用度体现为未因任何备件短缺而停机的设备数所占的

百分比。备件管理人员往往更加关注系统的可用度,而并不是单项备件的满足程度。下文将给出考虑备件需求相关的可用度计算模型。由于维修人员是否采用串件拼修策略会直接影响到由于备件短缺造成的系统停机数量,从而反作用于备件的需求,故将针对不串件拼修和串件拼修两种情况研究备件需求的动态特性,并给出可用度计算公式。为了便于描述系统状态变化,定义函数:

$$\varphi(x, y) = \begin{cases} x - y, & x > y \\ 0, & x \leq y \end{cases} \quad (3)$$

(1) 不串件拼修策略

在不执行串件拼修策略的情况下,系统的任意部件缺货均会造成系统停机,即同一台设备上最多有一个部件失效,所以各备件短缺总量不会超过系统数,因为当备件短缺数总量为 K 时,所有的系统均已停机,将不会产生备件需求,即任意时刻工作系统数量与备件短缺量之间的关系如下

$$WE = K - \sum_{k=1}^K BO_k \quad (4)$$

由于工作系统数量 $WE \geq 0$, 根据式(4) 可以得出

$$\sum_{k=1}^K BO_k \leq K \quad (5)$$

由于当 $DI_l \leq s_l$ 时, $BO_l = 0$, 当 $DI_l > s_l$ 时, $BO_l =$

$$q_{ij}^{(1)} = \begin{cases} (K - \sum_{h=1}^L \varphi(DI_h^{(i)}, s_h)) n_l \lambda_l, \\ \mu_l DI_l^{(i)}, \\ - \sum_{l=1, DI_l^{(R^+(i,l))} \neq 0}^L (K - \sum_{h=1}^L \varphi(DI_h^{(i)}, s_h)) n_l \lambda_l - \sum_{l=1, DI_l^{(R^-(i,l))} \neq 0}^L \mu_l DI_l^{(i)}, & i = j \\ 0, & \text{else} \end{cases}$$

由于连续不可约马尔可夫系统的稳态分布必然存在,记不考虑串件拼修的平稳分布为 $\pi_1 = [\pi_1^{(1)}, \pi_1^{(2)}, \dots, \pi_1^{(1J_1)}]$, 则有

$$\begin{cases} \pi_1 Q_1 = 0 \\ \sum_{i=1}^{1J_1} \pi_1^{(i)} = 1 \end{cases} \quad (9)$$

由式(9) 可以求解出稳态分布 $\pi_1 = [\pi_1^{(1)}, \pi_1^{(2)}, \dots, \pi_1^{(1J_1)}]$ 。不执行串件拼修策略时,对于备件配置方案 S , 系统的稳态可用度为

$$A(S) = \frac{K - \sum_{i=1}^{1J_1} \sum_{l=1}^L \pi_1^{(i)} \varphi(DI_l^{(i)}, s_l)}{K} \quad (10)$$

(2) 串件拼修策略

考虑串件拼修的情况,即将所有的备件空缺集中于停机设备上,记 $\lceil X \rceil$ 为不小于 X 的最小整数,所以任意时刻工作设备数量与备件短缺量之

$DI_l - s_l$, 所以有 BO_l 与 DI_l 的关系如式(6):

$$BO_l = \varphi(DI_l, s_l), \quad l = 1, 2, \dots, L \quad (6)$$

所以备件在修数满足如下关系

$$\sum_{l=1}^L \varphi(DI_l, s_l) \leq K \quad (7)$$

定义满足式(7) 的所有部件在修数组组合成了系统的可行状态集,记为 J_1 , $|J_1|$ 表示集合 J_1 中元素的个数。集合 J_1 中第 j 种部件在修数组组合 $(DI_1^{(j)}, DI_2^{(j)}, \dots, DI_k^{(j)})$ 为系统的第 j 种状态,记为状态 $DI^{(j)}$ 。 $R^+(j, l)$ 为状态 $DI^{(j)}$ 的第 l 种部件在修数 $DI_l^{(j)}$ 加 1, 而其他部件在修数不变所代表的状态编号,如果该状态不存在,则记 $R^+(j, l) = 0$; $R^-(j, l)$ 为状态 $DI^{(j)}$ 的第 l 种部件在修数 $DI_l^{(j)}$ 减 1, 而其他部件在修数不变所代表状态的编号,如果该状态不存在,则记 $R^-(j, l) = 0$ 。

可修复系统部件的失效率与修复率均为常数,系统状态互通,且状态有限,其构成一个连续不可约马尔可夫链。记该连续不可约马尔可夫链的最小生成元矩阵为 $Q_1 = [q_{ij}^{(1)}]$, $i = 1, 2, \dots, |J_1|$, $j = 1, 2, \dots, |J_1|$ 。最小生成元矩阵 Q_1 的矩阵系数为

$$\begin{aligned} j &= DI_l^{(R^+(i,l))}, l = 1, 2, \dots, L \\ j &= DI_l^{(R^-(i,l))}, l = 1, 2, \dots, L \\ i &= j \\ &\text{else} \end{aligned} \quad (8)$$

间的关系如下

$$WE = K - \lceil \max(\frac{BO_1}{n_1}, \frac{BO_2}{n_2}, \dots, \frac{BO_L}{n_L}) \rceil \quad (11)$$

由于工作系统数量 $WE \geq 0$, 根据式(11) 可以得出

$$\lceil \max(\frac{BO_1}{n_1}, \frac{BO_2}{n_2}, \dots, \frac{BO_L}{n_L}) \rceil \leq K \quad (12)$$

对于部件 l , 当 $BO_l = (K-1)n_l + 1$, 即代表所有的设备上均存在部件 l 短缺的情况, 此时所有设备均已停机。由于假设设备的部件失效相互独立, 即不可能多个部件同时失效。对于部件 l , $BO_l = (K-1)n_l + 1$ 成立时, 则对于部件 l' , $l' \neq l$, 必有 $BO_{l'} < (K-1)n_{l'} + 1$ 。故有各部件的期望短缺量满足以下关系

$$\sum_{l=1}^L \varphi(BO_l, (K-1)n_l) \leq 1 \quad (13)$$

对于部件 l , $BO_l = \varphi(DI_l, s_l)$, 故根据式(13)

可以得到各备件在修数满足如下关系

$$\sum_{l=1}^L \varphi(DI_l, s_l + (K-1)n_l) \leq 1 \quad (14)$$

满足式(14)的所有部件在修数组构成了系统的可行状态集,记为 J_2 。| J_2 |, $DI^{(j)}, R^+(j, l), R^-(j, l)$ 定义同上文所述,可修复系统各部件

$$q_{ij}^{(2)} = \begin{cases} (K - \lceil \max(\frac{\varphi(DI_1^{(i)}, s_1)}{n_1}, \frac{\varphi(DI_2^{(i)}, s_2)}{n_2}, \dots, \frac{\varphi(DI_L^{(i)}, s_L)}{n_L}) \rceil) n_l \lambda_l, & j = DI_l^{(R^+(i, l))}, l = 1, 2, \dots, L \\ \mu_l DI_l^{(i)}, & j = DI_l^{(R^-(i, l))}, l = 1, 2, \dots, L \\ - \sum_{\substack{l=1, \\ DI_l^{(R^+(i, l))} \neq 0}}^L (K - \lceil \max(\frac{\varphi(DI_1^{(i)}, s_1)}{n_1}, \frac{\varphi(DI_2^{(i)}, s_2)}{n_2}, \dots, \frac{\varphi(DI_L^{(i)}, s_L)}{n_L}) \rceil) n_l \lambda_l - \sum_{\substack{l=1, \\ DI_l^{(R^-(i, l))} \neq 0}}^L \mu_l DI_l^{(i)}, & i = j \\ 0, & \text{else} \end{cases} \quad (15)$$

记执行串件拼修策略时可修复系统的稳态分布为 $\pi_2 = [\pi_2^{(1)}, \pi_2^{(2)}, \dots, \pi_2^{(|J_2|)}]$, 则有

$$\begin{cases} \pi_2 Q_2 = 0 \\ \sum_{i=1}^{|J_2|} \pi_2^{(i)} = 1 \end{cases} \quad (16)$$

求解式(16)可以求出稳态分布 π_2 。在考虑串件拼修时,对于备件配置方案 S ,此时系统的稳态可用度为

$$A(S) = 1 - \frac{1}{K} \sum_{i=1}^{|J_2|} \pi_2^{(i)} \lceil \max(\frac{\varphi(DI_1^{(i)}, s_1)}{n_1}, \frac{\varphi(DI_2^{(i)}, s_2)}{n_2}, \dots, \frac{\varphi(DI_L^{(i)}, s_L)}{n_L}) \rceil \quad (17)$$

1.4 优化模型与模型求解

上文分别给出了考虑备件需求相关的情况下执行串件拼修与不串件拼修策略时系统稳态可用度的计算方法,故可以构建备件优化配置模型如下:

$$\begin{aligned} \min \quad & C(S) = \sum_{l=1}^L c_l s_l \\ \text{s.t.} \quad & A(S) \geq \bar{A}, \quad s_l > 0, s_l \in \mathbf{Z}, l = 1, 2, \dots, L \end{aligned} \quad (18)$$

其中 $C(S)$ 为备件方案 S 的单位时间配置费用, c_l 为部件 l 的单位时间配置费用, \bar{A} 为系统可用度约束, \mathbf{Z} 为整数集。下面将给出利用边际分析方法求解模型的过程。

本文将选用边际分析法对构建的备件优化配置模型进行求解,边际效应分析是一种渐进的优化技术,用于分配短缺资源来获得最大效益。它可以被认为是通过对边际单元的效益和费用的权衡分析,以达到对有效资源的合理利用。备件边际效

在修数组状态变化构成一个不可约连续马尔可夫链。记该连续不可约马尔可夫链的最小生成元矩阵为 $Q_2 = [q_{ij}^{(2)}], i = 1, 2, \dots, |J_2|, j = 1, 2, \dots, |J_2|$ 。最小生成元矩阵 Q_2 的矩阵系数为

应分析算法是一种在一定约束条件下依次进行迭代,直到系统满足最终目标值为止的优化算法。在每一轮的迭代过程中,关键是要根据优化效益分析,确定当前最需要调整的控制变量,以最有效地提高边际效益^[1]。

利用边际分析方法求解备件配置方案时,需要确定合理的备件初始配置方案,因为直接以各备件配置量为零作为初始解开始迭代无疑会有较大计算量,下面将给出备件初始配置方案的计算方法。考虑其他备件的配置数量趋向于无穷,设备停机均由部件 l 引起的情况,记 $\pi_l^{(j)}(s_l)$ 为该情况下部件 l 的备件配置量为 s_l 时 $DI_l = j$ 的稳态概率。考虑不执行串件拼修的情况,部件 l 满足系统可用度约束的最小备件数为

$$s_l^{(\min)} = \min \{s \mid \frac{K - \sum_{j=0}^{s+K} \pi_l^{(j)}(s) \varphi(j, s)}{K} \geq \bar{A}\} \quad (19)$$

考虑执行串件拼修的情况,部件 l 满足系统可用度约束的最小备件数为

$$s_l^{(\min)} = \min \{s \mid \frac{K - \sum_{j=0}^{s+(K-1)n_l+1} \pi_l^{(j)}(s) \lceil \frac{\varphi(j, s)}{n_l} \rceil}{K} \geq \bar{A}\} \quad (20)$$

如果系统满足可用度约束,则必有 $s_l \geq s_l^{(\min)}$,故可以取部件 l 优化前的初始配置量为 $s_l^{(\min)}$,利用边际分析方法求解本文模型的步骤如下:

步骤1 确定系统中 L 类备件的初始配置量,即 $s_l = s_l^{(\min)}$,计算初始备件配置方案下的系统可用度,如果满足约束,则初始备件配置方案即为最优的备件配置方案,否则转到步骤2;

步骤2 确定需要调整的备件种类,即寻找

最大的边际增量效益 $\delta_l(s_l | S)$ 计算公式为

$$\delta_l(s_l | S) = \begin{cases} \frac{A_l^+(S) - A(S)}{c_l}, & A_l^+(S) \leq \bar{A} \\ \frac{\bar{A} - A(S)}{c_l}, & A_l^+(S) > \bar{A} \end{cases}$$

(21)

其中 $A_l^+(S)$ 表示在现有的备件配置方案下,部件 l 的备件配置数量增加 1 件后系统的可用度;

步骤 3 当 L 类备件的 $\delta_l(s_l | S)$ ($l = 1, 2, \dots, L$) 值都确定后,将 $\delta_l(s_l | S)$ 最大的部件所对应的备件数量加 1,即 $s_{l_o} = s_{l_o} + 1, l_o = \max\{l | \delta_l(s_l | S)\}$,进入步骤 4;

步骤 4 计算调整后的备件方案对应的系统可用度,满足系统可用度约束,则迭代完成,当前的备件配置方案即为最优的备件配置方案,否则,转入步骤 2。

2 案例分析

某维修保障机构负责保障某型设备 25 台,该

表 1 不执行串件拼修情况下边际效益分析过程

Tab. 1 Process of marginal analysis in the case of non-cannibalization

循环次数	s_1	s_2	$\delta_1(s_1 S)$	$\delta_2(s_2 S)$	最优备件追加对象	$A(S)$ (当前可用度)	$A(s_{l_o} + 1 S)$ (本次循环结束可用度)
1	2	4	0.0211	0.0156	X1	0.8717	0.8928
2	3	4	0.0151	0.0165	X2	0.8928	0.9126
3	3	5	0.0074	0.0062	X1	0.9126	0.9287

表 2 执行串件拼修情况下边际效益分析过程

Tab. 2 Process of marginal analysis in the case of cannibalization

循环次数	s_1	s_2	$\delta_1(s_1 S)$	$\delta_2(s_2 S)$	最优备件追加对象	$A(S)$ (当前可用度)	$A(s_{l_o} + 1 S)$ (本次循环结束可用度)
1	2	4	0.0092	0.0048	X1	0.9108	0.9281

图 2 给出了不执行串件拼修策略与执行串件拼修策略下系统各部件的备件配置量与系统可用度之间的关系,图 3 给出了不执行串件拼修策略与执行串件拼修策略下利用本文的模型求解的备件配置费用与可用度之间的关系。不难发现,在相同的备件配置方案下,系统在执行串件拼修策略情况下与不执行串件拼修策略的情况相比具有更高的可用度水平,尤其在备件配置量较少的情况下。串件拼修策略的执行可以较大幅度地提高系统可用度,降低系统备件配置费用,故对于备件配置费用较高且具备串件拼修条件的系统应尽可能执行串件拼修策略。

种设备由 X_1 、 X_2 两类部件构成, X_1 、 X_2 的工作失效时间和修复时间均服从指数分布,故障率 λ_1 、 λ_2 分别为 0.0004h^{-1} 、 0.0003h^{-1} ,修复率 μ_1 、 μ_2 分别为 0.0025h^{-1} 、 0.0025h^{-1} ,机用数 n_1 、 n_2 分别为 1, 2,其单位时间配置费用分别为 10 元/h 与 12 元/h。系统的可用度约束为 0.92。下面将利用本文的备件优化模型针对串件拼修和不串件拼修两种情况确定满足可用度约束的最优备件配置方案。

通过式(19)和式(20)可以计算出不串件拼修与串件拼修两种情况下系统的备件初始配置方案均为 $s_1^{(\min)} = 2, s_2^{(\min)} = 4$ 。利用边际分析法可以得出不执行串件拼修策略时满足备件短缺量约束的最优备件配置方案为 $s_1 = 4, s_2 = 5$,此时系统的可用度为 0.9281,系统备件配置费用为 10 元/h。利用边际分析法可以得出执行串件拼修策略时满足备件短缺量约束的最优备件配置方案为 $s_1 = 3, s_2 = 4$,此时系统的可用度为 0.9287,系统备件配置费用为 7.8 元/h。不串件拼修与串件拼修两种情况下边际分析过程分别如表 1、表 2 所示。

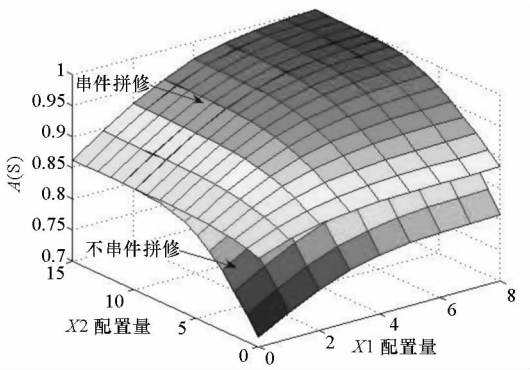


图 2 X_1 、 X_2 备件配置量与系统可用度关系
Fig. 2 Relation between spare allocation of X_1 , X_2 and system availability

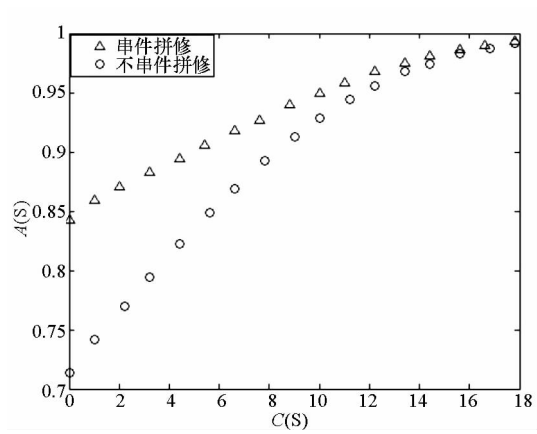


图3 系统总费用与系统可用度之间的关系
Fig.3 Relation between total cost and system availability

3 结 论

基于备件需求相互独立假设的备件需求预测往往会造成较大的误差,不能准确地预测系统实际的备件需求。本文以包含多台相同设备以及一个备件仓库的可修复库存系统作为研究对象,分析了系统中各种备件需求的相关性以及传统的基于备件需求相互独立的备件配置理论的不足之处。本文选取可用度作为系统的备件保障效能参数,然后利用连续马尔可夫链的相关理论,描述了系统各状态之间的转移规律,针对不串件拼修和串件拼修两种情况分别给出了系统稳态可用度的计算方法,然后以稳态可用度作为约束,以备件配置费用作为优化目标给出了备件优化配置模型以及模型的边际分析求解方法。最后通过案例分析,验证了该模型的可行性。本文在建模时将系统失效与维修假设为指数分布,对于工作寿命与维修时间不服从指数分布的情况,可以用 PH (phase-type distributions) 分布进行拟合,仍然可以按照本文的建模的基本思路构建优化模型,因为 PH 分布具有指数分布无记忆性的基本性质且任何分布总可以选择一个适当的 PH 分布把它拟合到任意精确的程度^[18]。

参考文献 (References)

[1] Sherbrooke C C. Optimal inventory modeling of systems; multi-echelon techniques[M]. 2nd ed. Boston: Kluwer Academic Publishers, 2004.

[2] Perlman Y, Mehrez A, Kaspi M. Setting expediting repair policy in a multi-echelon repairable item inventory system with limited repair capacity [J]. Journal of the Operational Research Society, 2001, (52): 198 - 209.

[3] Al-Rifai M H, Rossetti M D. An efficient heuristic optimization algorithm for a two-echelon (R, Q) inventory system [J]. International Journal of Production Economics, 2007, 109(1/2): 195 - 213.

[4] Wong H, van Houtum G J, Cattrysse D, et al. Multi-item

spare parts systems with lateral transshipments and waiting time constraints [J]. European Journal of Operational Research, 2006, 171(3): 1071 - 1093.

[5] Topan E, Bayındır Z P, Tan T. An exact solution procedure for multi-item two-echelon spare parts inventory control problem with batch ordering in the central warehouse [J]. Operations Research Letters, 2010, 38(5): 454 - 461.

[6] Caggiano K E, Jackson P L, Muckstadt J A, et al. Efficient computation of time-based customer service levels in a multi-item, multi-echelon supply chain: A practical approach for inventory optimization [J]. European Journal of Operational Research, 2009, 199(3): 744 - 749.

[7] de Smidt-Destombes K S, van der Heijden M C, van Harten A. On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy [J]. Reliability Engineering and System Safety, 2004, 83(3): 287 - 300.

[8] de Smidt-Destombes K S, van der Heijden M C, van Harten A. On the interaction between maintenance, spare part inventories and repair capacity for a k-out-of-N system with wear-out [J]. European Journal of Operational Research, 2006, 174(1): 182 - 200.

[9] de Smidt-Destombes K S, van der Heijden M C, van Harten A. Joint optimisation of spare part inventory, maintenance frequency and repair capacity for k-out-of-N systems [J]. International Journal of Production Economics, 2009, 118(1): 260 - 268.

[10] Kaplan A J. Incorporating redundancy considerations into stockage models [J]. Naval Research Logistics, 1989, 36(5): 625 - 638.

[11] Balkhi Z T, Benkherouf L. On an inventory model for deteriorating items with stock dependent and time-varying demand rates [J]. Computers & Operations Research, 2004, 31(2): 223 - 240.

[12] Dye C Y, Chang H J, Teng J T. A deteriorating inventory model with time-varying demand and shortage-dependent partial backlogging [J]. European Journal of Operational Research, 2006, 172(2): 417 - 429.

[13] Sakaguchi M. Inventory model for an inventory system with time-varying demand rate [J]. International Journal of Production Economics, 2009, 122(1): 269 - 275.

[14] Zhou Y W. A multi-warehouse inventory model for items with time-varying demand and shortages [J]. Computers & Operations Research, 2003, 30(14): 2115 - 2134.

[15] Lau H C, Song H, See C T, et al. Evaluation of time-varying availability in multi-echelon spare parts systems with passivation [J]. European Journal of Operational Research, 2006, 170(1): 91 - 105.

[16] 陈童, 黄卓, 郭波. 基于 PH 分布和 MAP 的备件(s, S)库存模型 [J]. 系统工程理论与实践, 2009, 29(6): 107 - 114.

CHEN Tong, HUANG Zhuo, GUO Bo. Spare parts (s, S) inventories system based on Phase-type distribution and Markovian arrival process [J]. Systems Engineering—Theory & Practice, 2009, 29(6): 107 - 114. (in Chinese)

[17] 程海龙, 康锐, 韦艺, 等. 备件满足率与备件短缺量对比分析 [J]. 北京航空航天大学学报, 2008, 24(5): 580 - 583.

CHENG Hailong, KANG Rui, WEI Yi, et al. Comparison analysis of fill rate and backorder [J]. Journal of Beijing University of Aeronautics and Astronautics, 2008, 24(5): 580 - 583. (in Chinese)

[18] Asmussen S, Bladt M. Point processes with finite-dimensional conditional probabilities [J]. Stochastic Processes and Their Applications, 1999, 82(1): 127 - 142.

战争设计工程中专家思维收敛过程的建模与分析^{*}

王长春,陈俊良,陈超

(国防科学技术大学 信息系统工程重点实验室,湖南 长沙 410073)

摘要:专家意见能否达成共识是战争设计工程中的一个关键问题。首先从复杂理论视角对专家思维过程的适应性、聚集性、非线性进行分析,接着将专家群体划分为两类并且运用马尔科夫转移矩阵对研讨过程建模,提出并证明了两类专家群体思维收敛的充分必要条件,得到群体思维收敛不仅与研讨过程中专家之间的关系影响矩阵有关,而且与专家的初始意见有关。最后通过计算机仿真进一步探索了群体思维收敛过程的影响因素。

关键词:战争设计工程;关系矩阵;马尔科夫链;共识

中图分类号:E917, C934 **文献标志码:**A **文章编号:**1001-2486(2012)03-0074-06

Modeling and analysis of the experts thought convergence in war design engineering

WANG Changchun, CHEN Junliang, CHEN Chao

(Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China)

Abstract: Expert's opinion consensus is one of key problems in the War Design Engineering. Firstly, the characteristics of experts thought, such as adaptive, cluster and non-linear, are analyzed from the view of complex theory. According to their interactive relationship, experts are divided in to two categories, and the convergence process is modeled by Markov transmission matrix. Then, the necessary and sufficient conditions of reaching consensus for each category experts are presented and proved. It ends with the conclusion that the condition of experts thought convergence is not only relative to interactive relationship matrix, but also the primal opinion of expert. Finally, some propositions are verified by computer simulation.

Key words: war design engineering; relationship matrix; Markov chain; consensus

战争是打出来的,也是设计出来的^[1]。为应对战争中的各种复杂性,2005年沙基昌教授提出了战争设计工程方法,其中专家群体是一个重要组成部分。但是由于专家知识背景,对问题看法价值观不同,使得众多专家意见是否具有收敛性,或者说群体思维能否达成共识是一个值得研究的问题。

目前对群体思维问题研究主要集中在以下3个问题。(1)群体思维收敛的调控,调控方法主要分为两类:一类是改变决策者权重以达成共识,如,Ben-Arieh,Chen在综合前人研究的基础上^[2],提出决策者在不改变意见的条件下,通过改变相关决策者权重的调控过程;另一类是改变专家的评价意见形成共识,如顾基发,Herrera-Viedma,王丹力通过个体一致性指标识别出成员意见分歧程度,引导个体修改其观点,若群体一致性指标大于或等于给定阈值时,研讨过程结束^[3-5]。(2)群

体思维的演化动力学。用连续或离散变量表示群体中个体的意见,通过设计个体之间的交互规则,分析群体思维的演化规律。如Hegslmann基于个体有限理性的交互规则^[6],分析比较了算术平均、几何平均、调和平均3种迭代机制下群体观点的演化规律;Weisbuch基于邻居效应的交互规则^[7],分析了网络结构、个体理性阈值对群体思维演化的影响。(3)群体思维收敛条件。De Groot首先运用向量表示群体的初始意见^[8],用左乘随机矩阵表示群体思维的迭代、反馈过程,提出并证明了群体思维收敛的条件。随后,许多研究者对该模型进行了扩展^[9-13],如:Gilardoni对比研究了线性集结算子和对数集结算子对群体思维收敛的影响^[11],分析并证明这两种意见集结方法在理论上是等价的。Chatterjee进一步推导了迭代矩阵随时间改变时群体思维的收敛条件^[13]。国内对群体思维收敛条件的定量研究较少,如:董

^{*} 收稿日期:2011-09-25

基金项目:国家自然科学基金资助项目(71101149)

作者简介:王长春(1983—),男,江西吉安人,博士研究生,E-mail:wcc198324@163.com;

陈超(通信作者),男,讲师,博士,E-mail:changshacc@163.com

玉成证明了当专家数足够多的时候^[14],加权几何平均综合判断矩阵与加权算术平均综合判断矩阵依概率收敛到客观排序向量。尽管对群体思维收敛条件研究已有一些结论,但是这些理论忽略了群体初始意见对思维收敛的影响。

1 战争设计工程中专家研讨过程分析

战争设计工程中的专家群体研讨是战争设计专家团队针对战争设计的目标,围绕装备与战法策略进行的研究讨论,并对装备与战法策略设计的合理性与正确性进行的思辨性论证。专家群体研讨是群体成员知识、经验相互交流的过程,是思维相互碰撞、主意相互激发的过程,是群体思辨与论证的过程,也是群体成员相互学习的过程。在群体专家研讨过程中,涉及多个领域的专家,他们面对一个共同的主题时可能有不同的意见,为了达成共识,需要进行多轮迭代与反馈,如图1所示。

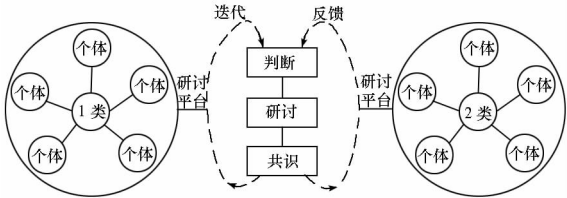


图1 专家群体共识形成过程

Fig.1 The consensus process of experts thought

专家群体研讨是一个动态系统,其复杂性主要表现在:

(1) 专家群体研讨的适应性

群体专家对战争问题的认识是一个不断改进、不断深化、旋进式收敛的过程。专家群体研讨的目的是达成共识,基于这样的假设,本文认为在研讨过程中专家会根据其他专家的意见来变换和调整自己的判断,并且得到群体思维收敛不仅与研讨过程中专家之间的关系影响矩阵有关,而且与专家的初始意见有关。

(2) 专家群体研讨的聚集性

聚集是复杂适应系统的一个基本特征,是指个体通过一定的条件,利用某种方式形成一个具有一定边界的更大个体的行为。一方面,专家通过某种组织形式聚集成为上层主体,上层主体又能够聚集成为更上层主体,组成复杂适应系统典型的层次结构。另一方面,专家在交互地学习和积累经验的过程中,各主体不断调整其结构和行为方式,持相同或相似意见的专家会形成小群体。本文用循环类来表示这种群体专家属性。

(3) 专家群体研讨的非线性

专家群体研讨的非线性是指决策主体的行为不是简单地叠加成为整个系统的行为,一个变量的微小变化对其他变量具有不成比例、甚至无法预测的影响。这主要体现在两个方面:首先是构成专家群体的成员要素,即决策主体的思想和决策能力,不具有可加性。其次,在专家群体研讨中通常都是由许多子系统组成,不同层次的子系统之间相互关联,相互制约,以某种或多种方式发生着复杂的非线性相互作用。

2 专家思维收敛过程模型

战争设计工程中专家研讨过程是一个动态的、非线性的演化过程。本文以专家群体对某一未知变量判断为背景(如:某型号导弹命中目标的概率),用专家的判断值来表示专家的思维,对其交互和收敛过程进行分析建模。为了明确研究问题,提出如下假设。

假设1 战争设计工程研讨的初始阶段,专家*i*的对某未知参数的判断为 $F_i^{(0)}$,*k*个专家群体的初始判断向量为 $\mathbf{F} = (F_1^{(0)}, F_2^{(0)}, \dots, F_k^{(0)})^T$ 。

假设2 第一轮研讨后,专家*i*的判断值调整为 $F_i^{(1)} = \sum_{j=1}^k a_{ij} F_j^{(0)}$,其中 a_{ij} 表示在研讨过程中专家*i*对专家*j*赋予的权重系数,可以用来反映专家*i*对专家*j*判断的信任和支持程度,且 $\sum_j a_{ij} = 1$ 。经*n*次研讨后,专家*i*判断值调整为: $F_i^{(n)} = \sum_{j=1}^k a_{ij} F_j^{(n-1)}$ 。

因此,战争设计工程中群体思维收敛过程可表示成为: $\mathbf{F}^{(n)} = \mathbf{A} \mathbf{F}^{(n-1)} = \mathbf{A}^n \mathbf{F}^{(0)}$,其中影响矩阵 $\mathbf{A} = (a_{ij})$, $\mathbf{A}^n = (a_{ij}^n)$, $i, j = 1, 2, \dots, k$,用来描述研讨过程中专家之间的影响关系,它由专家在研讨中的地位 and 专家观点的可信度来确定。

定义1 根据迭代过程中权重系数的性质来划分研讨过程中专家之间的关系。如果存在某个正整数 $n \geq 1$,使得 $a_{ij}^n > 0$,则在研讨过程中专家*j*的观点对专家*i*的判断有影响。同理,如果不存在 $n \geq 1$,使得 $a_{ij}^n > 0$ 成立,则在研讨过程中专家*j*的观点对专家*i*的判断没有影响。

定义2 如果研讨过程中群体中所有专家之间都存在着相互影响,那么该群体称为第一类专家群体;如果群体中存在着两个专家,在研讨过程中他们之间没有影响,那么该群体称为第二类专家群体。

定义3 随着研讨过程的推进,专家观点一直在反馈和调整,如果 $\mathbf{F}^{(n)} = \mathbf{F}^{(n-1)}$,那么专家群

体达成共识,或者说群体思维收敛。

根据以上定义,借鉴马尔科夫相关理论,容易得到第一类专家群体影响矩阵的周期 $d = 1$,且它是第二类专家群体的特例。为此我们重点分析第二类专家群体思维收敛过程。根据 Frobenius 理论,第二类专家群体的影响矩阵通过置换变换^[15],可以划分为 m 个循环类,每个循环类又可分为 d_i 个子类。其中, P_i 是 $m_i \times m_i$ 矩阵, P_{ij} 为 $m_{ij} \times m_{i(j+1)}$ 矩阵, d_i 为矩阵 P_{ij} 的周期。

$$A = \begin{bmatrix} P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & P_m \end{bmatrix}, \text{其中,}$$

$$P_i = \begin{bmatrix} 0 & P_{i1} & 0 & \cdots & 0 \\ 0 & 0 & P_{i2} & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & P_{id_i-1} \\ P_{id_i} & 0 & 0 & \cdots & 0 \end{bmatrix}。$$

对于 $i = 1, \dots, m, j = 1, \dots, d_i$, 令 A_{ij} 表示 $m_{ij} \times m_{ij}$ 矩阵, 且 $A_{ij} = P_{ij} \cdot P_{i(j+1)} \cdots P_{id_i} \cdot P_{i1} \cdots P_{i(j-1)}$, 那么

$$P_i^{d_i} = \begin{bmatrix} A_{i1} & 0 & \cdots & 0 \\ 0 & A_{i1} & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & 0 & A_{id_i} \end{bmatrix}$$

假设 $\pi(i, j) = (\pi(i, j)_1, \dots, \pi(i, j)_{m_{ij}})$ 为线性方程组 $\pi(i, j)A_{ij} = \pi(i, j)$, $\sum_{l=1}^{m_{ij}} \pi(i, j)_l = 1$ 的解。因为 P_{ij} 为不可约非周期矩阵, 所有 A_{ij} 也是不可约非周期矩阵, 进而得到线性方程组的解 $\pi(i, j)$ 存在且唯一。

3 专家思维收敛模型分析

3.1 专家思维收敛条件

命题 1: 对于第一类专家群体, 无论群体的初始观点是什么, 群体研讨最终都能够收敛。

证明: 首先证明第一类专家群体循环迭代 n 次后, 关系影响矩阵每一列的最小元素随着 n 的增加而增大, 每一列的最大元素随着 n 的增加而减小。

假设在循环迭代 n 次后的关系影响矩阵中, 第 j 列中的最小元素是 $\min_i a_{ij}^n = m_j(n)$, 第 j 列中的最大元素是 $\max_i a_{ij}^n = M_j(n)$ 。由切尔曼 - 科尔莫科洛甫方程, 得到

$$a_{ij}^n = \sum_k a_{ik} a_{kj}^{n-1} \geq \sum_k a_{ik} m_j(n-1) = m_j(n-1) \quad (1)$$

其中, $m_j(n) \geq m_j(n-1)$ 。同理可以得到

$$a_{ij}^n = \sum_k a_{ik} a_{kj}^{n-1} \leq \sum_k a_{ik} M_j(n-1) = M_j(n-1),$$

其中, $M_j(n) \leq M_j(n-1)$ 。

接着证明循环迭代 n 次后关系影响矩阵每一列的最小元素和最大元素, 随着 n 的增加趋于同一极限。即 $M_j(n), m_j(n)$ 趋于同一极限。运用马尔科夫链中的相关概念, 设从 i_0 经 n 步到达 j 的转移概率是 $m_j(n)$, 从 i_1 经 $n-1$ 步到达 j 的转移概率是 $M_j(n-1)$ 。则有

$$\begin{aligned} m_j(n) &= a_{i_0 j}^{(n)} = \sum_k a_{i_0 k} a_{kj}^{(n-1)} \\ &= \varepsilon a_{i_0 j}^{(n-1)} + (a_{i_0 i_1} - \varepsilon) a_{i_1 j}^{(n-1)} + \sum_{k \neq i_1} a_{i_0 k} a_{kj}^{(n-1)} \\ &\geq \varepsilon M_j(n-1) + [a_{i_0 i_1} - \varepsilon + \sum_{k \neq i_1} a_{i_0 k}] \cdot m_j(n-1) \end{aligned} \quad (2)$$

因此, $m_j(n) \geq \varepsilon M_j(n-1) + (1 - \varepsilon) m_j(n-1)$ 。

同理, 设从 i'_0 经 n 步到达 j 的转移概率是 $M_j(n)$, 从 i_2 经 $n-1$ 步到达 j 的转移概率是 $m_j(n-1)$, 则有

$$\begin{aligned} M_j(n) &= a_{i'_0 j}^n = \sum_k a_{i'_0 k} a_{kj}^{n-1} \\ &= \varepsilon a_{i'_0 j}^{n-1} + (a_{i'_0 i_2} - \varepsilon) a_{i_2 j}^{n-1} + \sum_{k \neq i_2} a_{i'_0 k} a_{kj}^{n-1} \\ &\leq \varepsilon m_j(n-1) + [a_{i'_0 i_2} - \varepsilon + \sum_{k \neq i_2} a_{i'_0 k}] \cdot M_j(n-1) \end{aligned} \quad (3)$$

所以, $M_j(n) \leq \varepsilon m_j(n-1) + (1 - \varepsilon) M_j(n-1)$, 由此得到

$$M_j(n) - m_j(n) \leq (1 - 2\varepsilon) [M_j(n-1) - m_j(n-1)] \leq (1 - 2\varepsilon)^{n-1}$$

进一步可以得到, 当 $n \rightarrow \infty, M_j(n) = m_j(n)$ 。这就证明了当群体为第一类专家群体时, 群体专家思维最终可以收敛。

命题 1 说明当专家群体为第一类群体时, 无论群体的初始判断差异有多大, 最终都能达成共识。这个结论一方面揭示了群体专家之间的影响关系对群体思维收敛过程起着十分关键的作用; 另一方面从理论上解释了目前通过识别共识度调控专家判断的方法肯定能够收敛。因为一旦发现判断有较大差异的专家, 要么调整其判断值, 要么降低其影响权值。

命题 2: 对于第二类专家群体, 群体思维收敛的充分必要条件是: 对于任意 $i = 1, \dots, m, j = 1, \dots, d_i$, 都有 $\pi(i, j)F(i, j) = F^*$ 。

必要性: 当专家群体达成共识时, 即对于群体

中任意一个专家 $l, \lim_{n \rightarrow \infty} F_l^{(n)} = F^*$ 都成立。假设 l 为群体第 i 循环类第 j 子类专家,根据上文定义可知 $\pi(i,j) = (\pi(i,j)_1, \dots, \pi(i,j)_{m_{ij}})$ 为线性方程组 $\pi(i,j)A_{ij} = \pi(i,j)$ 的解,故 $\lim_{n \rightarrow \infty} (P_l^{d_i})^{(n)} = \pi(i,j)$,进而 $\lim_{n \rightarrow \infty} F_l^{(nd_i)} = \lim_{n \rightarrow \infty} P_l^{(nd_i)} F = \pi(i,j)F(i,j)$,又因为 $\lim_{n \rightarrow \infty} F_l^{(nd_i)} = \lim_{n \rightarrow \infty} F_l^{(n)} = F^*$,故 $\pi(i,j)F(i,j) = F^*$ 。

充分性:对任意 $i = 1, \dots, m; j = 1, \dots, d_i$,满足 $\pi(i,j)F(i,j) = F^*$,下面要证明对于群体中任意一个专家 $l, \lim_{n \rightarrow \infty} F_l^{(n)}$ 存在且等于 F^* 。假设 l 为第 i 循环类第 j 子类的专家,由上节定义可以得到 $\lim_{n \rightarrow \infty} P_l^{(nd_i)}$ 存在。对于 $r = 0, \dots, d_{i-1}, \lim_{n \rightarrow \infty} P_l^{(nd_i+r)} = P_l^{(r)} \lim_{n \rightarrow \infty} P_l^{(nd_i)}$ 存在并且等于 $P_l^{(*)}(r) = (0\pi(i,j + r)\mathbf{0})$,进一步可以得到: $\lim_{n \rightarrow \infty} F_l^{(nd_i+r)} = \lim_{n \rightarrow \infty} P_l^{(nd_i+r)} F = \pi(i,j+r)F(i,j+r) = F^*$ 。这就证明了对于每一个 $\{nd_i + r\}$ 的子序列 $r = 0, \dots, d_{i-1}, \lim_{n \rightarrow \infty} F_l^{(nd_i+r)}$ 极限都存在且收敛于 F^* ,因此得到 $\lim_{n \rightarrow \infty} F_l^{(n)}$ 也收敛于 F^* 。

命题2 说明对于第二类专家群体,群体思维收敛的条件不仅与专家之间的影响矩阵有密切关系,而且与群体专家的初始意见有关。在实际研讨过程中,属于第二类专家群体的情况更加普遍。该结论启示我们在选择专家时,为了达成共识,不仅要考虑专家观点的相似性,而且需要考虑群体专家的适应性。如果某个专家太固执、或者学习能力较差,那么该群体研讨很难取得共识。

3.2 示例

下面通过一个例子来对上面的理论进行说明。假设由7个专家组成的第二类群体,研讨过程中的影响矩阵为 P ,且在研讨过程中关系矩阵不改变^[11]。根据 P 的属性,将专家群体划分为 P_1 和 P_2 两个循环类。

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

,运用上文的定义,第一个循环类 P_1 的相关参数为

第一个循环类 $P_1 = P_{11} = A_{11} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$,进一步可以通过求解方程组 $\Pi(1,1)A_{11} = \Pi(1,1)$,及 $\sum_1^3 \Pi(1,1)_l = 1$ 得到 $\Pi(1,1) = (4/11, 3/11, 4/11)$ 。

对于第二个循环类 $P_2 = \begin{bmatrix} 0 & P_{21} \\ P_{22} & 0 \end{bmatrix}$,其中 $P_{21} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$, $P_{22} = \begin{bmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{bmatrix}$,运用上文定义,得到: $A_{21} = \begin{bmatrix} 5/12 & 7/12 \\ 11/24 & 13/24 \end{bmatrix}$, $A_{22} = \begin{bmatrix} 1/3 & 2/3 \\ 3/8 & 5/8 \end{bmatrix}$ 。

与上述方法类似,通过求解方程组得到: $\Pi(2,1) = (11/25, 14/25), \Pi(2,2) = (9/25, 16/25)$ 。因此,该群体在研讨过程中,群体思维收敛的充要条件为:

$$4/14F_1 + 3/14F_2 + 7/14F_3 = 11/25F_4 + 14/25F_5 = 9/25F_6 + 16/25F_7$$

4 仿真实验及结果分析

将群体专家的意见用连续变量 F_i 来表示,专家之间的交互、迭代关系用马尔科夫转移矩阵 $A = (a_{ij})$ 来表示,那么群体专家的研讨过程中的思维收敛过程可以表示为 $F^{(m)} = A^{(m)}F^{(0)}$ 。其中, $a_{ij} = \begin{cases} 1/\#X(i,F), & \text{if } j \in X(i,F) \\ 0, & \text{otherwise} \end{cases}$, $X(i,F)$ 表示满足 $X(i,F) = \{ \|F_j - F_i\| \leq \varepsilon \}$ 所组成的专家集合, $\#X(i,F)$ 表示集合中专家的数量。运用Matlab仿真平台对群体思维的演化过程进行仿真。

(1) 影响矩阵对群体思维收敛影响

在专家研讨过程中,关系影响矩阵直接影响到思维收敛情况,本文在固定群体初始意见的前提下,通过调整仿真实验的输入参数,使得影响矩阵分别有3个循环类,2个循环类和1个循环类,其对应的仿真结果如图2所示。

图2的仿真结果显示,当关系影响矩阵有3个循环类时,群体专家研讨结束时有4种不同的观点;当影响矩阵有2个循环类时,群体专家研讨结束时有两种不同观点;当影响矩阵有1个循环类时,群体专家将达成共识。这组实验说明群体研讨过程中专家之间的影响矩阵对群体思维收敛具有十分重要的作用,且影响矩阵循环类越少,越容易

取得共识。

(2) 专家数量对思维收敛影响

关于专家数量对群体思维的影响仍然是一个悬而未决的问题。文献[13]认为在概率意义下,专家数量越多,群体越容易收敛。Hegselmann 认为专家数量对群体思维收敛的影响与专家初始意见有关^[5]。

笔者认为群体专家数量对思维演化的影响是间接的,其影响效果与增加新专家后关系影响矩

阵属性有关。如果新增加专家都比较自我,不愿意接受别人意见,那么群体专家数量越多,分歧越多;如果新增加的专家理解能力和影响力都比较大,那么增加专家个数有利于群体思维的收敛。为此,本节在运用相同初始意见分布的前提下,分别考察 100、400 个专家在影响矩阵有 2 个循环类和 4 个循环类 的情况下群体专家观点的演化特性,如图 3 所示。

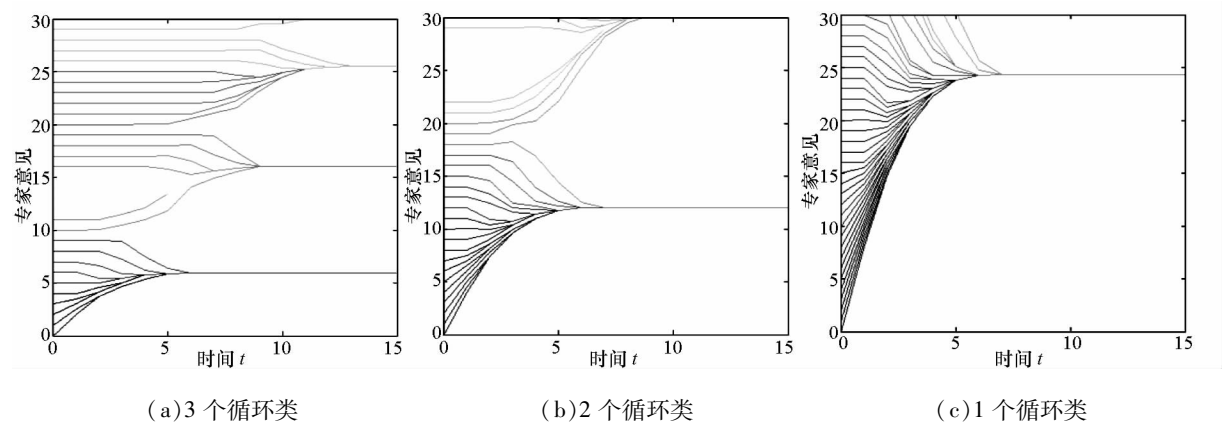


图 2 关系矩阵对群体思维收敛影响
Fig. 2 The effect of relationship matrix

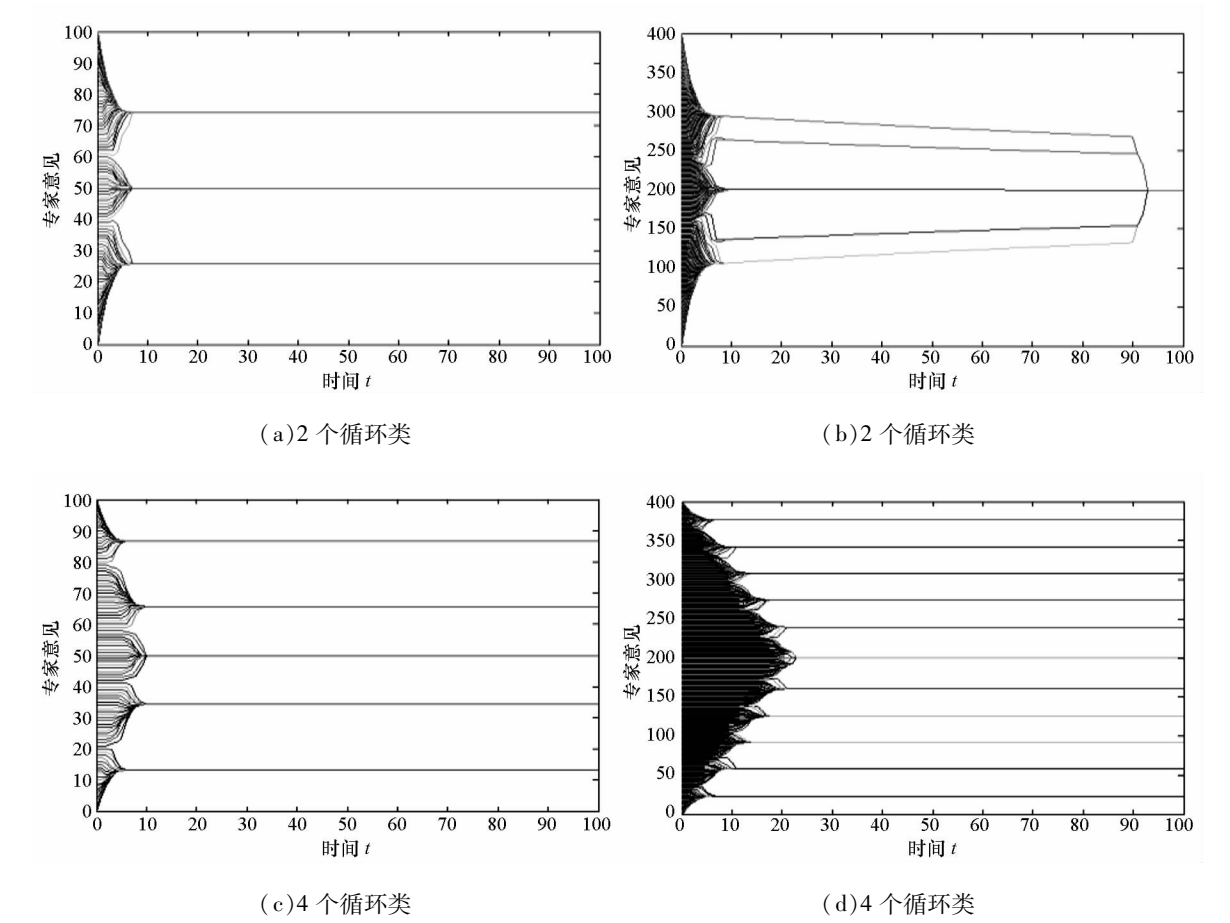


图 3 专家数量对群体思维收敛影响
Fig. 3 The effect of the expert number

图3(a)~(b)仿真结果显示,当影响矩阵只有2个循环类时,100个专家研讨结束时存在3种不同的意见,而当专家个数增加到400个时群体意见却可以收敛。图3(c)~(d)显示,当影响矩阵有4个循环类时,增加专家个数反而使得群体最终意见更加发散。这些结果表明专家群体数量对群体思维收敛的影响与关系影响矩阵的属性有关。

5 结 论

由于战争问题的复杂性和个体认知水平的有限性,群体专家经验和知识是有效解决问题的根本途径。本文将战争设计工程中专家研讨过程看成一个离散的动力系统,运用马尔科夫链对群体思维收敛过程进行建模,主要得到如下结论:

(1)提出了群体思维收敛的充分必要条件。群体思维收敛不仅与研讨过程中专家之间的关系影响矩阵有关,而且与专家的初始意见有关。

(2)运用仿真方法分析了关系影响矩阵、专家数量对群体思维收敛的影响。

本文试图从数学和计算机仿真两个角度推导和分析群体思维收敛的条件,相关结论对于战争设计过程中群体思维收敛问题具有一定的理论价值。

参考文献 (References)

[1] 沙基昌,毛赤龙,陈超. 战争设计工程 [M]. 北京: 科学出版社, 2009:74-92.
SHA Jichang, MAO Chulong, CHEN Chao. War design engineering [M]. Beijing: Science Press, 2009: 74-92. (in Chinese)

[2] Ben A D, Chen Z F. Linguistic group decision-making: opinion aggregation and measures of consensus [J]. Fuzzy Optimization and Decision Making, 2004, 5(4): 371-386.

[3] 顾基发. 意见综合——怎样达成共识 [J]. 系统工程学报, 2001, 16(5): 340-347.
GU Jifa. On synthesizing opinions how can we research consensus [J]. Journal of Systems Engineering, 2001, 16

(5): 340-347. (in Chinese)

[4] Herrera V E, Herrera F, Chiclana F. A consensus model for multiperson decision making with different preference structures [J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions, 2002, 32(3): 394-402.

[5] 王丹力,戴汝为. 专家群体思维收敛的研究[J]. 管理科学学报, 2002, 5(2): 1-5.
WANG Danli, DAI Ruwei. Reseach on convergence of expert group thought [J]. Journal of Mangement Sciences in China, 2002, 5(2): 1-5. (in Chinese)

[6] Hegselmann R, Krause U. Opinion dynamics and bounded confidence models, analysis, and simulation [J]. Journal of Artificial Societies and Social Simulation, 2002, 5(3): 1-33.

[7] Weisbuch G, Deffuant G, Amblard F. Meet, discuss, and segregate [J]. Complexity, 2002, 7(3): 55-63.

[8] De Groot M H. Reaching a consensus [J]. Journal of the American Statistical Association, 1974, 69(345): 118-121.

[9] Urbig D, Lorenz J, Herzberg H. Opinion dynamics: The effect of the number of peers met at once [J]. Journal of Artificial Societies and Social Simulation, 2008, 11(24): 1-27.

[10] Rauhut H, Lorenz J. The wisdom of crowds in one mind: how individuals can simulate the knowledge of diverse societies to reach better decisions [J]. Journal of Mathematical Psychology, 2011, 55(2):191-197.

[11] Berger R L. A necessary and sufficient condition for reaching a consensus using Degroot's method [J]. Journal of the American Statistical Association, 1981, 76: 415-418.

[12] Gilardon G L, Clayton M K. On reaching a consensus using Degroot's iterative pooling [J]. The Annals of Statistics, 1993, 21(1): 391-401.

[13] Chatterjee S, Seneta E. Toward consensus: some convergence theorems on repeated average [J]. Applied Probability, 1977, 89-97.

[14] 董玉成,徐寅峰,张桂清. 群体思维收敛性定量验证[J]. 系统工程理论与实践, 2006, 3: 108-111.
DONG Yucheng, XU Yinfeng, ZHANG Guiqing. On convergence of expert group thought [J]. Systems Engineering Theory & Practice, 2006, 3: 108-111. (in Chinese)

[15] 胡迪鹤. 随机过程论基础理论应用[M]. 武汉:武汉大学出版社,2000:271-326.
HU Dihe. The theory and application of stochastic process [M]. Wuhan: Wuhan University Press, 2000: 271-326. (in Chinese)

基于对偶四元数的单目视觉/惯性组合导航算法*

冯国虎, 吴文启

(国防科技大学 机电工程与自动化学院, 湖南 长沙 410073)

摘要: 提出一种适用于结构化道路的单目视觉/惯性组合导航定位算法。针对点特征匹配和连续多帧追踪受车速和相机视野制约的不足,提取道路上车道线的直线特征,引入对偶四元数描述直线特征。在基于对偶四元数的相对位姿估计算法的基础上,推导了图像特征增量与相机位姿增量的表达式。通过配准和时间同步,用惯导系统和相机分别解算的载体速度之差作为组合导航的观测量,建立 kalman 滤波修正组合导航系统的误差,包括相机测速标度因数误差。车载实验结果表明在结构化道路上算法是有效的。

关键词: 结构化道路;单目视觉;SINS;对偶四元数;组合导航

中图分类号: U666.1 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0080-05

Algorithm for monocular camera/SINS integrated navigation based-on dual quaternion

FENG Guohu, WU Wenqi

(College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China)

Abstract: An algorithm of monocular camera/SINS integrated navigation is presented for structural road. In order to overcome the difficulties of point features matching and tracking, which are restricted by carrier velocity and camera view, line features are extracted from the lane marks on the road. The line features are represented by dual quaternion. Compared with the relative pose estimation algorithm based on dual quaternion, the proposed algorithm deduced the formula of image feature increment and camera position increment. After registration and time synchronization, the velocity computation difference between the SINS and the camera was chosen as observation of integrated navigation. A Kalman filter was used to correct the integrated navigation error including the camera scale factor error of velocity measurement. The experiment results show that the proposed algorithm is accurate for structural road.

Key words: structural road; monocular camera; SINS; dual quaternion; integrated navigation

在具有清晰的车道线和道路导向箭头等交通标识的结构化道路上,准确地估计车辆运动参数对于自动驾驶和基于视觉的辅助驾驶起着非常重要的作用。在描述坐标系转换的数学工具中,对偶四元数能够以最简洁的形式表示一般性刚体运动。Goddard^[1]利用对偶四元数描述观测目标直线特征来确定目标相对相机坐标系的位置、姿态,该算法将位姿计算统一到对偶四元数体系中,对于确定运动目标的位姿是一种很好的思路。丁尚文等^[2]提出了基于对偶四元数的航天器相对位姿的双目视觉测量方法。张世杰^[3]系统研究了航天器相对导航中基于单目视觉测量目标航天器相对位姿的理论和算法。武元新^[4]研究了对偶四元数在捷联惯导系统中的两个基本理论问题:导航信息的数学描述和数值解算;导航信息的最优实时融合策略。

尽管视觉导航具有诸多优势,但仅使用视觉导航存在作用范围小,数据更新率低等不足,Olson^[5]证明了在仅有视觉信息的情况下,随着姿态误差的累积,长距离导航时系统的定位误差随距离呈非线性迅速增长,而惯性导航具备不受空间限制和采样频率高等优点,因此视觉/惯性组合导航在计算机视觉、微型飞行器自主导航、无人机自主着陆、空中加油和水下机器人等领域得到了广泛应用。Allen^[6]提出了用于无人机控制的基于视觉的惯性导航方法,利用相机从已知位姿信息的目标飞行器获得的导航参数,与惯性传感器测量的导航参数融合,得到修正的导航参数。Andreas^[7]提出了把惯性速率测量和基于一个特征信息的单目视觉测量相融合的算法。Mourikis等^[8]提出基于多位置极点约束单目视觉/惯性组合导航算法。

* 收稿日期:2011-11-12
基金项目:高等学校博士学科点专项科研基金资助项目(20069998009);教育部新世纪优秀人才支持计划(NCET-07-0225)
作者简介:冯国虎(1981—),男,湖南长沙人,博士研究生,E-mail:yantai_fgh@126.com;
吴文启(通信作者),男,教授,博士,博士生导师,E-mail:wengiwu_lit@sina.com

在结构化道路上,通常假设在非常短的时间内(相机拍摄周期),路面符合平坦假设,即路面为分段平面^[9-10]。在基于对偶四元数的视觉相对导航算法中,研究的大多是目标的相对位姿,目标运动,相机静止,研究载体的位姿,即相机随载体运动,目标静止的情况^[1-3]。在视觉/惯性组合导航算法中,相机或观测的是已知位姿信息的目标^[6],实际导航中往往难以满足这一约束,或观测点特征,使用多位置极点约束^[8]。点特征匹配尤其是连续多帧追踪同名点受车速和相机视野的制约,使用直线特征可以克服该不足。因为图像直线特征并不要求对应空间直线的同一段,直线的不同部位具有同一数学表达式。因此,本文采用直线特征,提取道路上车道线的直线特征,引入对偶四元数描述目标线特征,推导图像特征增量与相机位置增量的关系。

1 基于对偶四元数的单目视觉测距

选取北-东-地(NED)地理坐标系为导航坐标系,记为 n 系。如图1所示,相机安装在车体上,镜头朝前,相机坐标系 z 轴沿光轴方向,焦距为 f 。惯导系统安装在车内,惯导系统至相机的杆臂为 r^b 。

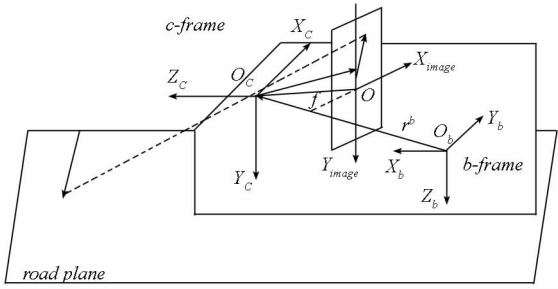


图1 单目视觉/惯性组合导航示意图

Fig. 1 Structure chart of monocular camera/INS integrated navigation

图像特征选取目标直线在像平面投影的垂点(line-point),根据Goddard^[1]的推导,垂点坐标为

$$x_{lp} = f \frac{m_x m_z}{m_x^2 + m_y^2}, y_{lp} = f \frac{m_y m_z}{m_x^2 + m_y^2} \quad (1)$$

相机运动可用对偶四元数算子描述^[4]

$$\hat{l}(\tau_2) = \hat{q}^*(\tau_1, \tau_2) \hat{l}(\tau_1) \hat{q}(\tau_1, \tau_2) \quad (2)$$

对偶部分展开,四元数联乘写成矩阵形式

$$m(\tau_2) = \bar{M}_{\Delta q} \bar{M}_{\Delta q}^+ m(\tau_1) + \frac{1}{2} \bar{M}_{\Delta q} \bar{M}_{\Delta q}^+ (\bar{M}_{\Delta t} + \bar{M}_{\Delta t}^+) l(\tau_1) \quad (3)$$

对偶部分增量为

$$\Delta m(\tau_1, \tau_2) = m(\tau_2) - m(\tau_1) = (M_q - I)m(\tau_1) + \frac{1}{2} M_q M_l l(\tau_1) \quad (4)$$

式中 $M_q = \bar{M}_{\Delta q} \bar{M}_{\Delta q}^+$, $M_l = \bar{M}_{\Delta t} + \bar{M}_{\Delta t}^+$ 。 M_q 和 M_l 分别是拍摄周期内相机姿态增量矩阵和位置增量矩阵。

由式(1),得

$$\begin{cases} \Delta x_{lp}(\tau_1, \tau_2) = \frac{\partial x_{lp}}{\partial m_x} \Delta m_x(\tau_1, \tau_2) + \frac{\partial x_{lp}}{\partial m_y} \Delta m_y(\tau_1, \tau_2) \\ \quad + \frac{\partial x_{lp}}{\partial m_z} \Delta m_z(\tau_1, \tau_2) \\ \Delta y_{lp}(\tau_1, \tau_2) = \frac{\partial y_{lp}}{\partial m_x} \Delta m_x(\tau_1, \tau_2) + \frac{\partial y_{lp}}{\partial m_y} \Delta m_y(\tau_1, \tau_2) \\ \quad + \frac{\partial y_{lp}}{\partial m_z} \Delta m_z(\tau_1, \tau_2) \end{cases} \quad (5)$$

在基于对偶四元数的相对位姿估计的传统方法中,得到的是图像特征与目标相对位姿的关系^[1-2]。本文推导的是图像特征增量与相机位姿增量的关系。传统方法中 $m(\tau_1)$ 和 $m(\tau_2)$ 是目标直线分别在目标系和相机系的对偶部分, M_l 是目标相对相机的位置矩阵。而本文 $m(\tau_1)$ 和 $m(\tau_2)$ 是目标直线在相机系中分别对应 τ_1 和 τ_2 时刻的对偶部分, M_l 是相机的位置增量矩阵。因此,传统方法要得到 $\Delta m(\tau_1, \tau_2)$ 与 $\Delta t_x, \Delta t_y, \Delta t_z$ 的关系,需要分别求 t_x, t_y, t_z 的偏导数。而本文 M_l 本身是 $\Delta t_x, \Delta t_y, \Delta t_z$ 的一阶表达式,只需将 $\Delta t_x, \Delta t_y, \Delta t_z$ 展开即可。 M_l 可以写成

$$M_l = \frac{\partial M_l}{\partial \Delta t_x} \Delta t_x + \frac{\partial M_l}{\partial \Delta t_y} \Delta t_y + \frac{\partial M_l}{\partial \Delta t_z} \Delta t_z \quad (6)$$

代入式(4),得

$$\begin{aligned} \Delta m(\tau_1, \tau_2) &= (M_q - I)m(\tau_1) + \frac{1}{2} M_q \frac{\partial M_l}{\partial \Delta t_x} l(\tau_1) \Delta t_x \\ &\quad + \frac{1}{2} M_q \frac{\partial M_l}{\partial \Delta t_y} l(\tau_1) \Delta t_y + \frac{1}{2} M_q \frac{\partial M_l}{\partial \Delta t_z} l(\tau_1) \Delta t_z \end{aligned} \quad (7)$$

代入式(5),写成矩阵形式

$$Z = \begin{bmatrix} \Delta x_{lp} \\ \Delta y_{lp} \\ \vdots \\ \Delta x_{nlp} \\ \Delta y_{nlp} \end{bmatrix} = H \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \end{bmatrix} + L = \begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix} \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \end{bmatrix} + \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix} \quad (8)$$

根据最小二乘,相机在拍摄周期内的位置增量为

$$[\Delta t_x, \Delta t_y, \Delta t_z]^T = (H^T H)^{-1} (H^T Z - H^T L) \quad (9)$$

根据路面为分段平面的假设^[9-10],相机在运动平面上的位置增量为

$$[\Delta t_x, \Delta t_z]^T = (H'^T H')^{-1} (H'^T Z - H'^T L) \quad (10)$$

式中

$$H' = \begin{bmatrix} H'_1 \\ \vdots \\ H'_n \end{bmatrix}, H'_i = \begin{bmatrix} H_{i11} & H_{i13} \\ H_{i21} & H_{i23} \end{bmatrix} \quad (11)$$

为方便求解, M_q 用欧拉角表示。假设载体在 (τ_1, τ_2) 内的姿态角增量为小角度, 有

$$\det(H'^T H') = \frac{f^4 y_{ilp}^3 y_{jlp}^3 l_{ix} l_{jx} (l_{ix} l_{jx} - l_{ix} l_{iz})^2 (y_{ilp} y_{jlp} l_{ix} l_{jx} + a \Delta \theta)}{h^4 ((f_{ix} + x_{ilp} l_{iz})^2 + y_{ilp}^2 l_{iz}^2) ((f_{jx} + x_{jlp} l_{jz})^2 + y_{jlp}^2 l_{jz}^2)} \quad (12)$$

式中

$$a = 2l_{ix} l_{jz} x_{ilp} y_{ilp} + 2l_{iz} l_{jx} x_{ilp} y_{jlp} + 2f_{ix} l_{jx} (y_{ilp} + y_{jlp}) \quad (13)$$

$l_i(\tau_1) = [l_{ix}, l_{iy}, l_{iz}]^T, l_j(\tau_1) = [l_{jx}, l_{jy}, l_{jz}]^T, \Delta \theta$ 为

载体在 (τ_1, τ_2) 内的俯仰角增量。
当提取的直线有两条满足
 $y_{ilp} y_{jlp} \neq 0, l_{ix} l_{jx} \neq 0, l_{ix} l_{jz} - l_{jx} l_{iz} \neq 0$ (14)
 $\det(H'^T H') \neq 0$ 成立, 方程有唯一解。

2 速度模式的组合导航

2.1 系统方程

速度模式的组合导航如图 2 所示。状态量定义为

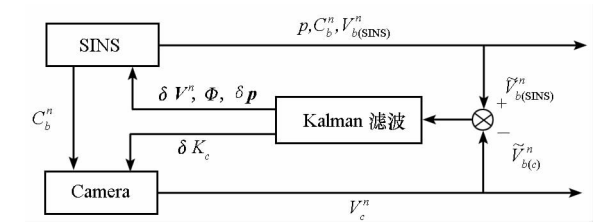


图 2 速度模式的组合导航示意图
Fig. 2 Structure chart of velocity integrate form

$$X(t) = [\delta V^n, \Phi^T, \delta p^T, \delta K_c]^T \quad (15)$$

式中 $\delta V^n = [\delta V_N, \delta V_E]^T, \Phi = [\varphi_N, \varphi_E, \varphi_D]^T$ 和 $\delta p = [\delta \lambda, \delta L]^T$ 分别是惯导系统解算的速度误差, 姿态误差和经、纬度误差; δK_c 是相机测速标度因数误差。

相机误差有安装偏差角和测速标度因数误差 δK_c 。认为安装偏差角是固定值, 预先标校, 不作为误差状态。测速标度因数 $K_c = h/f, h$ 为相机安装高度, 与路面状况有关, 为缓慢变化量。 δK_c 包括常值误差和随机误差, 常值误差当作不随时间

变化的随机常数处理, 随机误差用一阶马尔科夫过程表示, 反时间相关常数为 $\frac{-1}{\tau_c}$ 。

状态方程

$$\dot{X}(t) = \begin{bmatrix} F_{SINS} & \mathbf{0}_{7 \times 1} \\ \mathbf{0}_{1 \times 7} & \frac{-1}{\tau_c} \end{bmatrix} X(t) + W \quad (16)$$

式中 F_{SINS} 的表达式详见文献[11]。

2.2 观测方程

惯导系统和相机分别解算的载体速度测量值为

$$\tilde{V}_{b(SINS)}^n = V_{b(SINS)}^n + \delta V_{b(SINS)}^n \quad (17)$$

$$\tilde{V}_{b(c)}^n = V_{b(c)}^n + \delta V_{b(c)}^n = V_{b(c)}^n + V_{b(c)}^n \times \Phi + V_{c_e}^n \delta K_c \quad (18)$$

$$\text{式中 } V_b^n = C_b^n C_c^b V_c^n, V_c^n = \frac{\Delta t(\tau_1, \tau_2)}{\Delta \tau}, \Delta \tau = \tau_2 - \tau_1 \text{ 为相机拍摄周期。}$$

相机系和载体系之间的转移矩阵 C_b^c 为

$$C_b^c = T_b^{c*} T_b^{c*} \quad (19)$$

式中 T_b^{c*} 为载体系到相机系的转移矩阵, 由两者坐标系定义不同造成。 T_b^{b*} 为相机的安装偏差角矩阵, 当偏差角很小 ($< 3^\circ$) 时, 可以用差分旋转矩阵 (differential rotation matrix) 表示安装偏差角矩阵^[12]。

如图 1 所示, 惯导系统到相机的杆臂为 r^b , 相机速度 V_c^n 和相机解算的载体速度 $V_{b(c)}^n$ 之间存在杆臂效应, 两者的关系为

$$V_{b(c)}^n = V_c^n - C_b^n [\omega_{nb}^b \times r^b] \quad (20)$$

认为 r^b 和安装偏差角是固定值, 采用 Bäumker 等人的离线标定方法获得^[12]。

取惯导系统和相机分别解算的载体速度在北向和东向的差值为量测向量。系统的量测方程为

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} (\tilde{V}_{b(SINS)}^n - \tilde{V}_{b(c)}^n) = H_x X(t) + v(t) \quad (21)$$

观测矩阵

$$H_x = \begin{bmatrix} 1 & 0 & 0 & V_{b(c)D}^n & V_{b(c)E}^n & 0 & 0 & -V_{cN}^n \\ 0 & 1 & -V_{b(c)D}^n & 0 & V_{b(c)N}^n & 0 & 0 & -V_{cE}^n \end{bmatrix} \quad (22)$$

式(21)中, $V_{b(c)N}^n, V_{b(c)E}^n$ 和 $V_{b(c)D}^n$ 分别表示相机解算的载体速度在北向、东向和地量的分量; V_{cN}^n 和 V_{cE}^n 分别表示相机速度在北向和东向的投影; $v(t)$ 为量测噪声。

3 实验验证

相机和惯导系统装载在汽车上作车载实验,

如图 1 所示,相机安装在车体上,水平朝前拍摄。相关实验参数:陀螺零偏稳定性 $0.01^{\circ}/h(1\sigma)$,加速度计零偏稳定性 $5 \times 10^{-5} g(1\sigma)$,采样频率为 200Hz ,姿态更新频率为 100Hz 。相机分辨率 768×576 ,帧率 20f/s ,焦距 25mm ,像素尺寸 $9.9\mu\text{m}$,像点定位误差 $6\mu\text{m}$,相机安装高度 1.17m 。

采用 Hough 变换提取道路上车道线的直线特征。图 3 描述了某帧图像目标直线及其垂点的提取结果,图中粗线为提取的目标直线,垂点用“*”标记。

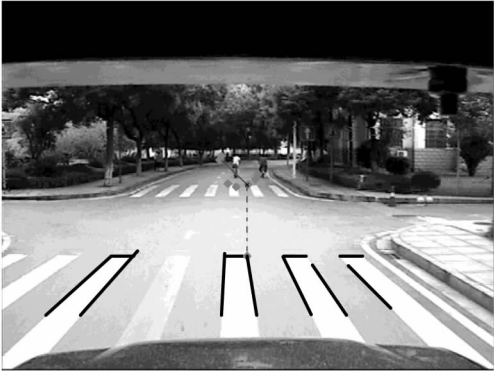


图 3 目标直线和垂点的提取

Fig. 3 The line features and line-point were extracted from a camera image

本文方法利用了惯性导航的优势,不用相机估计自身姿态,惯导系统连续提供拍摄时刻对应的相机三维姿态。同时,在图像信息缺失期间,采用惯导系统单独导航,确保导航的连续性。

本文方法估计的车载运动轨迹如图 4 所示。相对于起点,运动终点的估计值为北向 26m ,东向 -10m 。经测量,真实值为北向 0m ,东向 10m 。运动距离为 3.2km ,定位误差为 1.03% 。图 5 给出了本文方法估计的位置、姿态和速度精度。

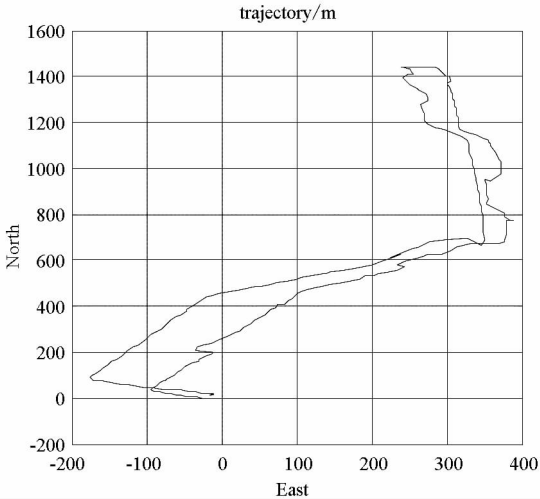


图 4 本文方法估计的车载运动轨迹

Fig. 4 The estimated trajectory of the proposed algorithm

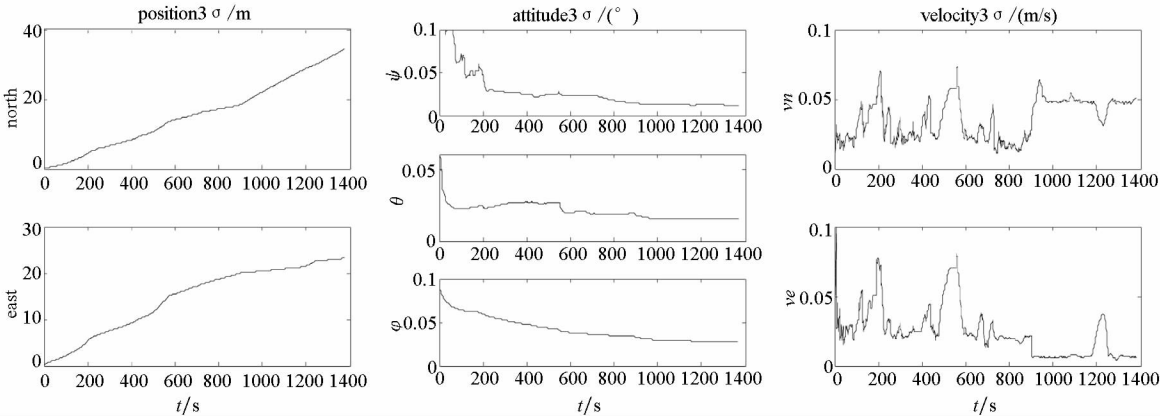


图 5 本文方法的位置、姿态和速度精度

Fig. 5 The 3σ bounds for the errors in the position, attitude and velocity of the proposed algorithm

研究单目视觉/惯性组合导航的文献中^[8],图像特征选取的是点特征,运动距离为 3.2km ,定位误差为 0.3% 。本文方法和文献中的方法在硬件条件上不一致,如惯性器件,摄像设备以及车辆行驶条件都不相同,可比性不大。不过,在硬件条件一致,组合导航方式相同(速度组合模式)的前提下,基于点特征的单目视觉/惯性组合导航算法,运动距离为 1.4km ,定位误差为 4.0% ^[13],相比之下,本文方法,运动距离为 3.2km ,定位误差为 1.03% ,具有一定的优越性。

4 结 论

为克服点特征匹配和连续多帧追踪同名点受车速和相机视野制约的不足,提取道路上车道线的直线特征,引入对偶四元数描述直线特征。利用对偶四元数算子构造相机位置增量的表达式,将相机估计的速度与惯导系统测量的速度进行组合,建立滤波方程来修正导航参数。车载实验结果表明本文方法直线特征匹配正确率高,能够连续多帧准确跟踪,车辆导航定位精度较高。

参考文献 (References)

[1] Goddard J S. Pose and motion estimation from vision using dual quaternion-based extended kalman filtering [D]. Knoxville: the University of Tennessee, 1997.

[2] 丁尚文, 王惠南, 刘海颖, 等. 基于对偶四元数的航天器交会对接位姿视觉测量算法 [J]. 宇航学报, 2009, 30 (6): 2145 – 2150.

DING Shangwen, WANG Huinan, LIU Haiying, et al . Algorithm of vision measure for relative position and pose of RVD spacecrafts based on dual-quaternion [J]. Journal of Astronautics, 2009, 30 (6): 2145 – 2150. (in Chinese)

[3] 张世杰. 基于单目视觉的航天器相对导航理论与算法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2005.

ZHANG Shijie. Study on theory and algorithms of monocular vision-based navigation for spacecraft [D]. Harbin: Harbin Institute of Technology. (in Chinese)

[4] 武元新. 对偶四元数导航算法与非线性高斯滤波研究 [D]. 长沙: 国防科技大学, 2005.

WU Yuanxin. Research on dual-quaternion navigation algorithm and monocular gaussian filtering [D]. Changsha: National University of Defense Technology, 2005. (in Chinese)

[5] Olson C F, Matthies L H, Schoppers M, et al. Rover navigation using stereo ego-motion [J]. Robotics and Autonomous Systems, 2003, 43 (4): 215 – 229.

[6] Allen D W, Johnson N, Alison A P. Vision-aided inertial navigation for flight control [J]. Journal of Aerospace Computing, Information and Communication, 2005, (2): 348 – 360.

[7] Andreas H. Relative position sensing by fusing monocular vision and inertial rate sensors [D]. Stanford: Stanford University, 2003.

[8] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation [C]// Proceedings of the IEEE International Conference on Robotics and Automation, Roma, Italy, 2007.

[9] Stein G P, Mano O, Shashua A. A robust method for computing vehicle ego-motion [C]// Proceedings of the IEEE Intelligent Vehicles Symposium, Dearborn, USA, 2000.

[10] 刘威, 魏存伟, 赵逢, 等. 基于单目视觉的自行车运动参数鲁棒估计 [J]. 机器人, 2009, 31 (1): 20 – 26.

LIU Wei, WEI Cunwei, ZHAO Feng, et al. A robust method for vehicle ego-motion estimation based on monocular vision [J]. Robot, 2009, 31 (1): 20 – 26. (in Chinese)

[11] 李万里. 车载组合导航自适应滤波及抗野值算法研究 [D]. 长沙: 国防科技大学, 2008.

LI Wanli. Research on adaptive Kalman filter and fault-tolerant algorithm used in in-vehicle integrated navigation system [D]. Changsha: National University of Defense Technology, 2008. (in Chinese)

[12] Bäumker M, Heimes F J. New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of an hybrid inertial navigation system [M]. Frankfurt am Main: Bundesamt für Kartographie und Geodäsie, 2002: 197 – 212.

[13] Feng G H, Wu W Q, Cao J L, et al. Algorithm for monocular visual odometry/SINS integrated navigation [J]. Journal of Chinese Inertial Technology, 2011, 19 (3): 210 – 215.

[10] Pousi J. Decision analytical approach to effect based operations [D]. Espoo: Helsinki University of Technology, 2009.

[11] Yaman D, Polat S. A fuzzy cognitive map approach for effect based operations: an illustrative case [J]. Information Sciences, 2009: 382 – 403.

[12] Wilkins D E, Desimone R V. Applying an AI planner to military operations planning [R]. SRI International Research Report, 1993.

[13] Myers K, Tyson W M. PASSAT: A user-centric planning framework [C]// Proceedings of the 3rd International NASA Workshop on Planning and Scheduling for Space. Houston, USA. 2002.

[14] Tate A, Levine J, Jarvis P, et al. Using AI planning technology for army small unit operations [C]// Proceedings of the Fifth International Conference on Artificial Intelligence Planning Systems, 2000.

[15] Potter S, Tate A, Wickler G. Using I-X process panels as intelligent to do lists for agent coordination in emergency response [C]// Proceedings of the Information Systems for Crisis Response and Management, Newark, New Jersey, USA, 2006.

[16] Fudenberg D, Tirole J. Game theory [M]. Massachusetts: The MIT Press, 1991.

[17] John H. Games with incomplete information played by “ bayesian ” players [J]. Management Science, 1967: 159 – 182.

(上接第 67 页)

太阳辐射作用下的飞机表面温场分布实验研究*

张春华,赵元,喻煌超,陈循
(国防科技大学装备综合保障技术重点实验室,湖南长沙 410073)

摘要:提出一种通过缩比模型测试飞机在太阳辐射作用下的表面温度场分布,进而通过有限元方法实现太阳辐射温场的三维模拟显示的实验方法。该方法避免了外场测试条件中自然日光辐射强度不可控和环境温度变化的影响,为飞机在太阳辐射作用下的表面温场理论模型验证和实测提供一种有效的方法,对飞机红外特性、日历寿命等研究具有借鉴意义。

关键词:太阳辐射;飞机表面温场;温场分布;三维模拟显示
中图分类号:TH133.1 **文献标志码:**A **文章编号:**1001-2486(2012)03-0085-04

Experimental research on surface temperature field of airplane under solar radiation

ZHANG Chunhua, ZHAO Yuan, YU Huangchao, CHEN Xun

(Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China)

Abstract: This paper presents an experimental method to measure surface temperature field distribution of airplane under solar radiation by airplane downsizing model, and suggests a method to realize analogue display of 3 dimensions for temperature field distribution by finite element method. This method may avoid the influence of the uncontrollable intensity of natural solar radiation and the variable ambient temperature under field measuring condition, and provide an efficient method to validate the theoretical model of surface temperature field distribution of airplane under solar radiation and to measure it by experiment. It will be beneficial to the research of infrared characteristics and calendar life for airplane.

Key words: solar radiation; surface temperature field of airplane; temperature field distribution; analogue display of 3 dimensions

飞机在停机坪停放时,太阳辐射作用的红外热效应会在机体蒙皮结构表面产生一定的附加温场分布,并且沿飞机表面形成完整的红外图像。目前红外制导导弹多采用8~14μm波段的红外成像制导,而飞机蒙皮热辐射能量主要集中于这个波段,因此飞机蒙皮温场是形成飞机红外图像的基本因素。随着红外探测性能的提高,将可能基于太阳辐射热效应产生的附加温场实现机型识别。此外,由于太阳辐射在机体蒙皮表面造成温度分布梯度,这种温度梯度的长期累积作用将造成飞机蒙皮结构的损伤,从而影响飞机的日历寿命。因此,飞机蒙皮在太阳辐射作用下的表面温场分布特性已成为飞机设计、研制、使用中的重要问题。

文献[1~3]研究了飞机蒙皮红外辐射热效应的理论模型和模拟计算方法,指导飞机设计对红外特性进行理论计算和数值模拟,但是这些模型和数值方法的准确性有待进一步的实验验证。文献[4]介绍了应用动态温度测试系统测量发动

机进口瞬态温场的方法,通过对温度场数据的修正、处理及计算,绘制温场波形图,但不能直观反映出温度场的三维分布情况。

一般情况下,在外场实际条件下对太阳辐射造成的温场进行实测是一种可行的技术途径,文献[5~6]研究了太阳能板上温场分布的相关方法;文献[7]研究了采用红外热成像测试建筑物外表面温场的方法,但是实测的方法往往受到外场测试条件中自然日光辐射强度不可控和环境温度变化的影响,难以实现飞机表面温场的准确测量。为了克服这一问题,在实验室通过太阳辐射设备和飞机缩比模型进行表面温场测试成为一种有效的替代途径。该方法可以准确模拟控制日光辐射强度和环境温度,实现温场准确测量,并且可以应用于在飞机设计阶段对飞机设计方案的太阳辐射特性进行实测研究。

针对上述问题,本文提出一种通过缩比模型测试飞机在太阳辐射作用下的表面温度场分布的实验方法,避免了外场测试条件中自然日光辐射

* 收稿日期:2012-01-06
基金项目:国防科技大学本科生创新实验资助项目(20090301)
作者简介:张春华(1974—),男,重庆人,副教授,博士。E-mail:chzhang@nudt.edu.cn

强度不可控和环境温度变化的影响,并通过有限元方法实现太阳辐射温场的三维模拟显示,为飞机在太阳辐射作用下的表面温场理论模型验证和实测提供一种有效的方法,对于飞机红外特性、日历寿命等研究具有借鉴意义。

1 实验方案

本文采用的实验方案如图 1 所示,利用太阳辐射试验设备模拟自然太阳光对飞机蒙皮的红外加热效应,在蒙皮表面建立温度场测试传感器阵列,温度传感阵列将温度信号转换为电信号,由数据采集器进行数据采集和记录。其中,太阳辐射设备的辐射强度为 $1120\text{W}/\text{m}^2$,温度范围 $-10\sim 120^\circ\text{C}$,满足 GJB150.7《军用设备环境试验方法 太阳辐射试验》的技术要求;飞机模型为 XX 高仿金属缩比模型,可以较好地反映飞机表面温场分布的实际情况;温场测试系统由多路数据采集接口、多路开关、传感器探头及引线组成,可测量温度范围为 $-200\sim +630^\circ\text{C}$,分辨率为 0.01°C 。

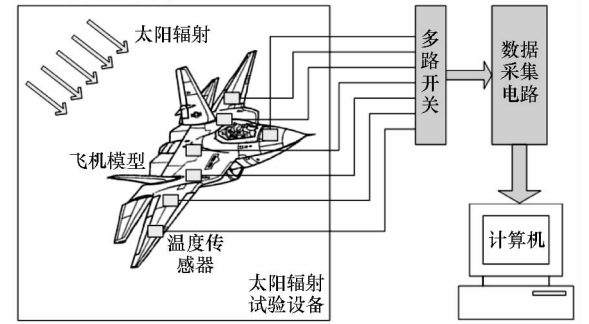


图 1 太阳辐射作用下的飞机表面温场测试方案
Fig. 1 Experimental scheme to measure surface temperature field of airplane under solar radiation

2 飞机蒙皮表面温场测试实验

在有限测点的情况下,综合考虑蒙皮形状的变化梯度和可能存在的温度变化梯度进行测点布局优化,以减小表面温场测试和拟合误差。考虑到表面温场的对称性,在模型单面选定了 14 个测量点,如图 2 所示。温度传感器感应部分通过锡箔包裹后固定在模型表面,避免太阳辐射对传感器直接加热。本实验采用 KEITHLEY 2700 组建温场测试系统,将所选测点传感器信息采集反馈,并通过环采通数据采集软件实现自动温度巡检,对温场测试设备按需要进行配置,对采集的温场数据进行实时记录。

按照 GJB150.7 的试验条件对飞机缩比模型

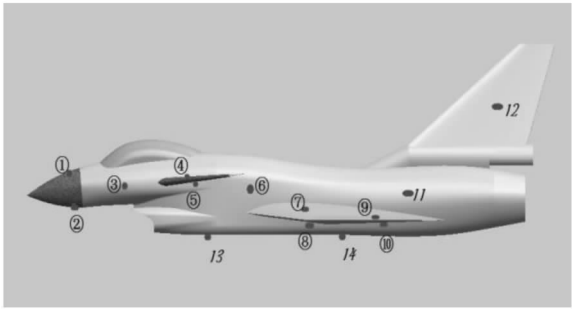


图 2 温度测点在飞机模型表面的分布情况
Fig. 2 Temperature measuring point distribution on the surface of airplane model

加载太阳辐射,总辐射强度为 $1120(\pm 10\%) \text{W}/\text{m}^2$,此条件代表了我国大部分地区的自然太阳辐射条件。在具体加载太阳辐射条件时还应满足以下要求:①飞机缩比模型在太阳辐射试验设备内的放置应保证气流不受阻挡;②太阳辐射光源离开飞机缩比模型充足的距离,尽可能在飞机缩比模型范围内模拟太阳光的平行照射。在此条件下,分别在环境温度为 20°C 、 25°C 、 30°C 、 35°C 、 40°C 情况下测试蒙皮表面温场数据,测试实验场景如图 3 所示。实验流程如下:

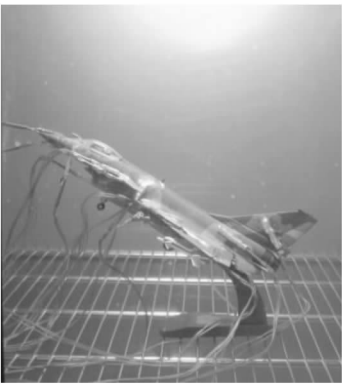


图 3 温度测试实验场景
Fig. 3 Temperature measuring experiment

- ①预处理:将实验用的飞机模型表面清洁干净,并在正常的实验大气条件下放置 1h;
- ②在开始对飞机缩比模型加载太阳辐射条件的同时使温度测试系统开始工作;
- ③分布在飞机蒙皮表面的温度传感器阵列测量出各个布点的温度,通过多路数据接口和温度采集系统软件对数据进行实时采集和记录;
- ④在工作界面中显示数据实时变化,当温场数据达到稳定时停止加载太阳辐射条件,结束数据采集。表 1 为测得的太阳辐射作用下飞机表面各测量点的稳态温度数据。

表 1 太阳辐射作用下飞机表面各测量点的稳态温度

Tab.1 Stable temperature of measuring points on surface of airplane under solar radiation

测点 编号	环境温度(℃)				
	20	25	30	35	40
1	28.66	32.47	38.92	43.87	48.5
2	29.21	33.95	39.48	44.25	48.23
3	34.01	38.12	44.06	48.86	52.1
4	32.81	36.6	43.04	47.96	53.2
5	29.79	34.72	40.14	45.35	50.22
6	36.43	40.21	46.64	51.61	54.38
7	40.05	42.65	49.88	54.37	57.01
8	36.69	40.98	46.87	51.58	53.92
9	42.15	44.64	52.02	56.59	58.64
10	34.67	38.87	44.78	49.74	52.52
11	39.63	42.86	49.54	54.31	56.5
12	29.48	34.03	40.04	44.74	47.78
13	29.11	33.64	39.16	44.15	47.47
14	31.01	35.49	41.42	46.39	49.25

3 飞机蒙皮表面温场拟合

3.1 飞机缩比模型三维造型

根据缩比模型实际测量尺寸,利用三维造型软件采用放样扫描再对称的方法进行飞机缩比模型三维造型,完成的缩比模型三维造型如图 4 所示。考虑到简化后续有限元网格划分及降低温度场分布拟合计算量的需要,对图 4 所示的模型进行简化,并在有限元分析软件中对简化飞机模型表面进行有限元网格划分,得到如图 5 所示的简化三维造型。

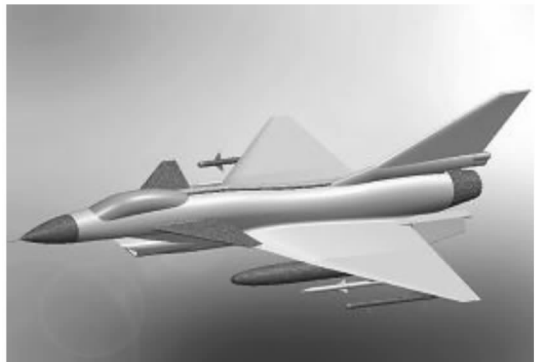


图 4 飞机缩比模型的三维造型
Fig.4 3-dimension model of the airplane in experiment

3.2 温度数据加载及温场云图显示

以表 1 所示的各节点测量温度作为温度边界

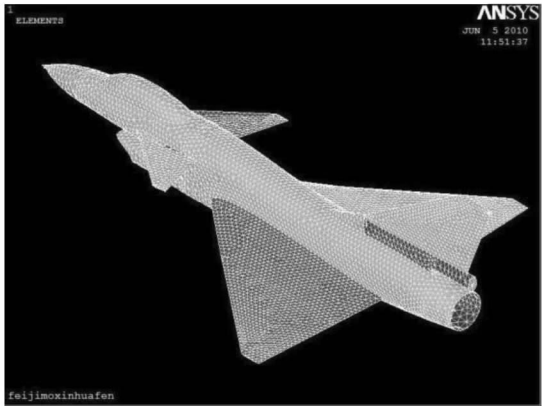


图 5 飞机缩比模型表面的网格划分
Fig.5 Grid division for surface of simplified airplane model in experiment

条件,在飞机模型上选择和试验测点相对的节点输入试验温度数据,以试验箱环境温度作为整个飞机模型的热载荷,利用有限元分析软件计算和云图显示功能直接显示飞机缩比模型表面的温度云图,显示不同环境温度条件下的太阳辐射在飞机表面造成的附加温场分布。图 6、图 7 分别为环境温度为 20℃ 和 25℃ 时飞机模型表面的温度云图。

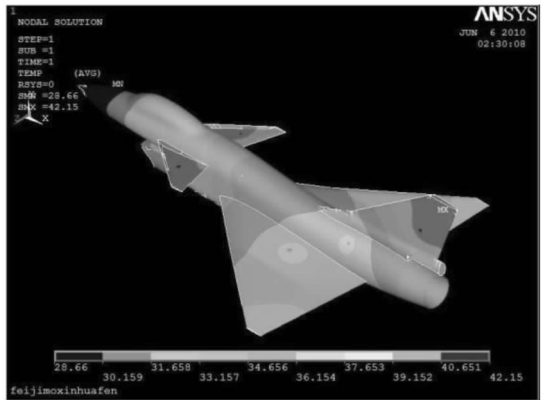


图 6 飞机模型表面的温度云图(环境温度为 20℃)
Fig.6 Temperature nephogram for surface of simplified airplane model in experiment (when ambient temperature is 20℃)

从图 6、图 7 的温度分布云图可以总结出飞机模型在太阳辐射条件下表面温场分布的规律:

- 1) 飞机模型表面温场呈阶梯状分布,机身头部和垂直尾翼受辐射面积小,温度较低,两侧机翼由于太阳辐射面积大,温度较高;
- 2) 随着环境温度的升高,飞机模型上的阶梯状温度分布规律基本保持一致。
- 3) 由于机翼较薄,热传导效率较高,所以温度分布比较均匀,机身材料复杂,热传导效率较低,所以温度分布梯度较大。

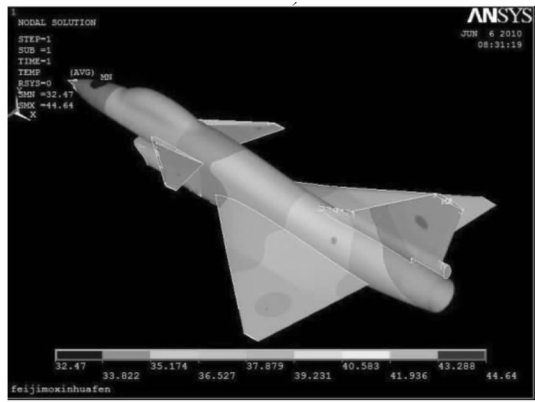


图 7 飞机模型表面的温度云图(环境温度为 25℃)

Fig.7 Temperature nephogram for surface of simplified airplane model in experiment (when ambient temperature is 25℃)

4 结 论

1) 本文为实际测试飞机蒙皮在太阳辐射作用下的表面温场分布提供了可行的实验方法,该方法排除了自然日光辐射强度不可控、环境温度不稳定等因素的干扰,并且借助有限元方法实现了飞机蒙皮温场分布的三维显示,可以直观表示出温场沿飞机蒙皮表面的分布情况;

2) 太阳辐射作用在飞机蒙皮表面造成的附加温场反映了飞机的基本外形特征,可以为飞机红外导引和反导、飞机红外隐身设计等研究提供依据;

3) 通过实验测定飞机蒙皮在太阳辐射作用下的温场分布情况,可以在此基础上进一步计算飞机蒙皮结构的热应力分布,对飞机蒙皮结构的日历寿命研究提供依据。

参考文献 (References)

[1] 艾青,夏新林,唐尧. 求解飞机蒙皮耦合热效应的壁面热流函数法[J]. 工程热物理学报, 2006, 27(4): 635 - 637. AI Qing, XIA Xinlin, TANG Yao. The wall heat flux function for solving coupled heat transfer of aircraft skin[J]. Journal of Engineering Thermophysics, 2006, 27(4): 635 - 637. (in Chinese)

[2] 夏新林,艾青,任德鹏. 飞机蒙皮红外辐射的瞬态温度场分析[J]. 红外与毫米波学报, 2007, 26(3): 174 - 177. XIA Xinlin, AI Qing, REN Depeng. Analysis on the transient temperature-fields for infrared radiation of aircraft skin[J]. Journal of Infrared and Millimeter Waves, 2007, 26(3): 174 - 177. (in Chinese)

[3] 翟泉慧,王强,额日其太. 高速飞机红外成像特性数值模拟研究[J]. 激光与红外, 2002, 32(3): 146 - 148. ZHAI Quanhui, WANG Qiang, Eriqitai. Numerical simulation of IR image of high-speed aircraft[J]. Laser and Infrared, 2002, 32(3): 146 - 148. (in Chinese)

[4] 齐亦农,李承辉,赵刚. 发动机进口瞬态温度场测试技术[J]. 航空发动机, 1999(3): 37 - 43. QI Yinong, LI Chenghui, ZHAO Gang. Testing technique of the transient temperature-fields in the engine entrance [J]. Aeroengine, 1999(3): 37 - 43. (in Chinese)

[5] Sadykov B S. Distribution of temperature field of a flat-plate solar collector and evaluation of its efficiency [J]. Applied Solar Energy, 1988, 24(5): 49 - 54.

[6] Avezov R R, Avezova N R, Samiev K A. Temperature field and heat flows through transparent covers of flat-plate solar collector housing[J]. Applied Solar Energy, 2008, 44(1): 13 - 16.

[7] 周克印,田裕鹏,姚恩涛,等. 建筑物温度场红外测试的研究[J]. 航空计测技术, 2004, 24(4): 5 - 8. ZHOU Keyin, TIAN Yupeng, YAO Entao, et al. Study of the method for measuring the temperature field of the building by infrared technology [J]. Aviation Metrology and Measurement Technology, 2004, 24(4): 5 - 8. (in Chinese)

基于 RCS 的三维低可探测性轨迹优化方法^{*}

陈 璟, 陈少飞, 刘鸿福

(国防科技大学 机电工程与自动化学院, 湖南 长沙 410073)

摘 要: 针对基于雷达散射截面(RCS)规避雷达威胁的飞行轨迹优化问题,提出了低可探测性三维轨迹优化的求解方法。通过B样条拟合构建连续可微的RCS数据模型,结合三维飞行动力学模型,建立规避雷达威胁下的飞行运动控制模型。将轨迹优化问题描述成为最优控制问题,其中飞行姿态控制、轨迹约束、边界条件作为约束条件,以降低雷达探测概率和减少飞行时间为目标函数。运用高斯伪谱法(GPM)将连续的最优控制问题转换为离散的非线性规划问题进行求解。仿真结果证明本文方法实现了求解单基地雷达和双基地雷达探测环境中低可探测性三维轨迹优化问题,有效降低了飞行过程中的雷达探测概率和暴露时间。

关键词: 姿态控制;低可探测性;雷达散射截面;三维轨迹优化;Gauss 伪谱法
中图分类号: V249.12 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0089-05

A Three-Dimensional low observable trajectory optimization method based on RCS

CHEN Jing, CHEN Shaofei, LIU Hongfu

(College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China)

Abstract: The method of finding an optimal aircraft trajectory to minimize the probability of detection by opponent radar detection systems is proposed. RCS data model, a three dimension (3D) aircraft and detection model, was established first. And then, the problem was described as an optimal control problem. The constraint functions were constructed by attitude control, trajectory constraints and boundary conditions. The cost function was formed to minimize both the probability of detection and the total flight time. After that, the continuous optimal control problem was parameterized to discrete nonlinear programming problem by Gauss pseudospectral method (GPM). Simulation results demonstrate the feasibility of the proposed method in low observable three-dimension trajectory optimization problem with monostatic radar and bistatic radar system, and it is also shown that the probability of detection and the exposure time decrease significantly.

Key words: attitude control; low observable; radar cross section; 3D trajectory planning; Gauss pseudospectral method

近年来,世界军事强国都在加强各自的空中作战力量,特别加紧对高性能、低可探测性飞行器的研制和作战应用研究,基于雷达探测模型的飞行轨迹优化问题成为新的研究热点^[1]。逐步开展了建立基于雷达散射截面积(Radar Cross Section, RCS)特征的雷达探测约束模型^[2]以及无人机低可探测性轨迹生成方法^[3]的相关研究。

目前,针对微小型飞行器低可探测性飞行轨迹规划,大多数研究停留在航迹规划层面,有些对飞行器目标特性进行了简化,或者没有将航迹规划和姿态控制综合考虑^[4-5]。现代飞行器 RCS 值的普遍降低和与姿态相关的目标特性,使得飞行器在飞行过程中处于暴露与不暴露状态的频繁切换之间,常规的目标特征模型不足以刻画现代飞行器的目标特性。飞行器 RCS 与飞行姿态(如方位角、俯仰角)的紧密耦合,使得结合飞行姿态

控制的轨迹优化方法成为求解飞行器低可探测性轨迹优化问题新的研究思路。相关研究通过建立姿态调整规划代价,对低 RCS 微小卫星姿态实时规划的方法进行了探索^[6]。基于 RCS 姿态特性的轨迹优化,通过底层的飞行器机动特性和姿态控制,求解出用于飞行的相关状态量和控制量,具有较强的应用价值。

作为直接法求解最优控制问题的一种方法,高斯伪谱法(Guass Pseudospectral Method, GPM)采用全局插值多项式的有限基在一系列离散点上近似状态变量和控制变量,因其较少的参数和较高的精度优势,被认为具有应用于实际系统实时最优控制问题的潜力^[7]。

本文结合飞行动力学模型,通过B样条拟合方法对RCS数据进行拟合建模;结合雷达探测模型建立了飞行运动控制模型。建立了低可探测性

^{*} 收稿日期:2011-09-25
作者简介:陈璟(1972-),男,江西南昌人,副教授,硕士生导师,E-mail:Chenjing001@vip.sina.com

轨迹优化的相关约束和目标函数,将问题描述成为最优控制问题。通过 Gauss 伪谱法将问题转换为离散的非线性规划问题,由大规模非线性规划求解器 SNOPT^[8] 计算得出飞行轨迹。通过仿真实验得到双基地雷达系统威胁环境中的轨迹优化结果。

1 模型建立

1.1 飞行器动力学模型

由于俯仰角和方位角的微小改变都会引起飞行器 RCS 的强烈变化,以及飞行器姿态与雷达探测概率的紧密耦合关系,需要选用三维的飞行动力学模型。在轨迹优化问题研究当中,通常选取如下的质心方程作为飞行动力学模型:

$$\begin{cases} m\dot{v} = T\cos\alpha - D - mg\sin\gamma \\ m\dot{v}\dot{\gamma} = T\sin\alpha\cos\varphi + L\cos\varphi - mg\cos\gamma \\ m\dot{v}\dot{\psi}\cos\gamma = T\sin\alpha\sin\varphi + L\sin\varphi \\ \dot{h} = v\sin\gamma \\ \dot{m}_f = -b \\ \dot{x} = v\cos\gamma\cos\psi \\ \dot{y} = v\cos\gamma\sin\psi \end{cases} \quad (1)$$

其中 m 为飞行器质量, v 为速度, α 为迎角, γ 为航迹倾斜角, ψ 为航迹方位角, φ 为滚转角; x 、 y 分别表示飞行器在东、北方向与原点的距离, h 为飞行高度; T 表示发动机的推力,并假设只与飞行的马赫数 M 和飞行高度 h 有关; D 和 L 分别为阻力和升力。这些变量中,迎角 α 和滚转角 φ 为控制量,其他均为状态量。

1.2 飞行器 RCS 特征模型

由于 Gauss 伪谱法将最优控制问题转换为非线性规划问题时,为了满足一阶最优条件要求,目标函数需要满足一阶连续可微,从而选用 B 样条拟合的方法构建 RCS 特征模型。

假设固定雷达频率下,飞行器 RCS 特征只与其相对雷达的方位角、俯仰角两个姿态角有关。文献[1]建立了飞行器 RCS 随机体相对雷达方位角、俯仰角变化的数据模型。本文对该 RCS 数据模型进行三次 B 样条函数拟合建模。其中,矩形区域上由其张量积网格控制生成的双三次 B 样条函数为^[9]

$$\sigma = f(\kappa_r, \theta_r) = \sum_{k=0}^3 \sum_{l=0}^3 B_k(\kappa_r) B_l(\theta_r) \delta_{(i+k)(j+l)} \quad (2)$$

飞行器 RCS 的模型均是按照不同方向照射飞行器时对应 RCS 取值的形式给出,若计算某一

飞行状态下飞行器的 RCS 值,首先需要得到当前状态飞行器在机体坐标系下雷达的照射方向。

假设 κ_r 和 θ_r 分别为飞行器相对于雷达的方位角和俯仰角, σ 为 RCS 值。为了计算某一飞行状态下 RCS 值,首先需要计算对应的 κ_r 和 θ_r 。通过坐标变换,由地面坐标系下飞行器相对于雷达方向向量转换机体坐标系再计算得出 κ_r 和 θ_r 。

地面坐标系下,设雷达坐标为 (x_r, y_r, z_r) 。则飞行器相对于雷达方向的单位向量为

$$\mathbf{x}_e = \begin{pmatrix} x - x_r \\ y - y_r \\ h - z_r \end{pmatrix} / \sqrt{(x - x_r)^2 + (y - y_r)^2 + (h - z_r)^2} \quad (3)$$

由地面坐标系到气流坐标系的转换矩阵为^[10]

$$\mathbf{S}_{we} = \begin{bmatrix} \cos\gamma\cos\psi & & \\ \sin\gamma\cos\psi\sin\varphi + \sin\psi\cos\varphi & & \\ \sin\gamma\cos\psi\cos\varphi - \sin\psi\sin\varphi & & \\ \cos\gamma\sin\psi & -\sin\gamma & \\ \sin\gamma\sin\psi\sin\varphi + \cos\psi\cos\varphi & \cos\gamma\sin\varphi & \\ \sin\gamma\sin\psi\cos\varphi - \cos\psi\sin\varphi & \cos\gamma\cos\varphi & \end{bmatrix} \quad (4)$$

由气流坐标系到机体坐标系的转换矩阵为

$$\mathbf{S}_{bw} = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix} \quad (5)$$

从而可以得出机体坐标系下飞行器相对于雷达方向的单位向量

$$\mathbf{x}_b = \mathbf{S}_{bw} \mathbf{S}_{we} \mathbf{x}_e \quad (6)$$

然后得到计算飞行器 RCS 的参数 κ_r 和 θ_r

$$\kappa_r = \arctan\left(\frac{x_{b,2}}{x_{b,1}}\right) \quad (7)$$

$$\theta_r = \arctan\left(\frac{x_{b,3}}{\sqrt{x_{b,1}^2 + x_{b,2}^2}}\right) \quad (8)$$

其中, $x_{b,i}$ 表示 \mathbf{x}_b 中第 i 个元素。

1.3 双基地 RCS 计算

双基地雷达对于目标的探测更为有效。对于收发分置的双基地雷达,根据由目标单基地 RCS 计算双基地 RCS 的经验公式,将双基地 RCS 表示为与目标的单基地 RCS 和双基地角相关的函数^[11],即

$$\sigma_{bi} = \sigma_{mon} [1 + \exp(|\beta|K - 2.4K - 1)] \quad (9)$$

其中, σ_{mon} 为目标单基地 RCS, β 为双基地角, K 为经验系数,由目标结构和复杂程度确定,且

$$K = \frac{\ln[4\pi A^2 / (\lambda^2 \sigma_{mon})]}{\pi - 2.4} \quad (10)$$

$$\beta = \arccos\left(\frac{R_1^2 + R_2^2 - b^2}{2R_1R_2}\right) \quad (11)$$

其中, A 表示在垂直于雷达波束方向上投影的目标面积, λ 为波长; R_1 、 R_2 分别表示目标到收站和发站的距离, b 为收、发站之间的距离。

1.4 雷达探测概率模型

设接收机的两个通道噪声服从高斯分布,则在幅度检波后进行目标探测时,噪声背景为瑞利分布。设目标信号也服从瑞利分布。文献[12]中推导出单个雷达在距离 R 处对目标的探测概率为

$$P_d = \exp\left(\frac{\ln(P_f)R^4}{K\sigma}\right) \quad (12)$$

其中 P_f 表示雷达系统的虚警概率; K 表示与雷达威力相关的规整化因子,包含了反映雷达目标探测威力的内在因素; σ 为 RCS 值。

从而,由 n 个雷达构成的组网雷达系统对飞行器的探测概率表示为

$$PD = 1 - \prod_{k=1}^n (1 - P_d(k)) \quad (13)$$

其中 $P_d(k)$ 表示第 k 个雷达对目标的探测概率。

2 基于 Gauss 伪谱法的轨迹优化方法

Gauss 伪谱法求解轨迹优化方法的思路,是将轨迹优化这样连续的最优控制问题描述成为离散的非线性规划问题进行求解。对于 Gauss 伪谱法求解最优控制问题的详细说明见文献[7],这里针对轨迹优化问题作简要介绍。

首先,通过线性变换,时间 t 可以表达为 τ 的函数

$$t = [(t_f - t_0)\tau + (t_f + t_0)]/2 \quad (14)$$

其中 t_0 是初始时刻, t_f 是结束时刻。从而满足标准的时间间隔 $\tau \in [-1, 1]$ 。

轨迹优化问题描述为在时间段 $[t_0, t_f]$ 上最小化目标函数

$$J = \varphi(\mathbf{x}(-1), t_0, \mathbf{x}(1), t_f) + \frac{t_f - t_0}{2} \int_{-1}^1 g[\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_0, t_f] d\tau \quad (15)$$

其中状态量 $\mathbf{x} \in R^{N_x}$, 控制量 $\mathbf{u} \in R^{N_u}$, 并满足以下约束

$$\frac{d\mathbf{x}}{d\tau} = \frac{t_f - t_0}{2} \mathbf{f}(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_0, t_f) \quad (16)$$

$$\psi(\mathbf{x}(-1), t_0, \mathbf{x}(1), t_f) = 0 \quad (17)$$

$$\mathbf{C}(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_0, t_f) \leq 0 \quad (18)$$

直接法求解轨迹优化是将问题离散化描述成为非线性规划问题。Gauss 伪谱法基于拉格朗日差值多项式来逼近状态量和控制量。由 N 次拉格

朗日差值多项式 L 逼近状态量

$$\mathbf{x}(\tau) \approx \mathbf{X}(\tau) = \sum_{i=0}^N \mathbf{X}(\tau_i) L_i(\tau) \quad (19)$$

$N-1$ 次拉格朗日差值多项式 L^* 逼近控制量

$$\mathbf{u}(\tau) \approx \mathbf{U}(\tau) = \sum_{i=0}^{N-1} \mathbf{U}(\tau_i) L_i^*(\tau) \quad (20)$$

从而将微分约束描述成代数约束:

$$\sum_{i=0}^N D_{ki} \mathbf{X}_i - \frac{t_f - t_0}{2} \mathbf{f}(\mathbf{X}_k, \mathbf{U}_k, \tau; t_0, t_f) = 0 \quad (21)$$

其中 $\mathbf{X}_k = \mathbf{X}(\tau_k)$, $\mathbf{U}_k = \mathbf{U}(\tau_k)$, $(k = 1, \dots, N)$ 。并且终端状态 $\mathbf{X}_f = \mathbf{X}(1)$ 满足 Gauss 积分公式

$$\mathbf{X}_f \equiv \mathbf{X}_0 + \frac{t_f - t_0}{2} \sum_{k=1}^N \omega_k \mathbf{f}(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) = 0 \quad (22)$$

通过 Gauss 积分公式对连续的目标函数进行逼近

$$J \equiv \varphi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) + \frac{t_f - t_0}{2} \sum_{k=1}^N \omega_k g(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \quad (23)$$

其中, ω_k 是 Gauss 权重。边界条件表示为

$$\psi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) = 0 \quad (24)$$

结合状态和控制变量信息的轨迹约束为

$$\mathbf{C}(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \leq 0 \quad (25)$$

通过以上过程实现将连续的轨迹优化问题描述成目标函数和代数约束的非线性规划问题,再通过相关的大规模非线性规划求解工具包进行求解。

3 基于 RCS 低可探测性轨迹优化问题求解

3.1 轨迹优化问题求解框架

基于 RCS 的低可探测性轨迹优化问题求解框架如图 1 所示。首先,通过对雷达探测相关数据进行 B 样条拟合,结合三维飞行性能模型和雷达的探测威力关系建立飞行运动控制模型。将飞行性能模型、雷达探测模型、航路点状态信息和其他输入信息综合构建约束条件和搜索空间。代价函数由飞行时间和雷达探测概率来综合表达。然后通过 Gauss 伪谱法将连续的最优控制问题转换为离散的非线性规划问题。最后通过非线性规划求解工具求解得到结果。

3.2 目标函数的建立

低可探测性轨迹优化是以飞行时间和探测概率为代价函数的动态优化问题。优化目标是找到综合考虑飞行时间最少和雷达探测概率最小,从起始点开始经过各个航路点到达终点的飞行轨迹。建立轨迹优化的目标函数

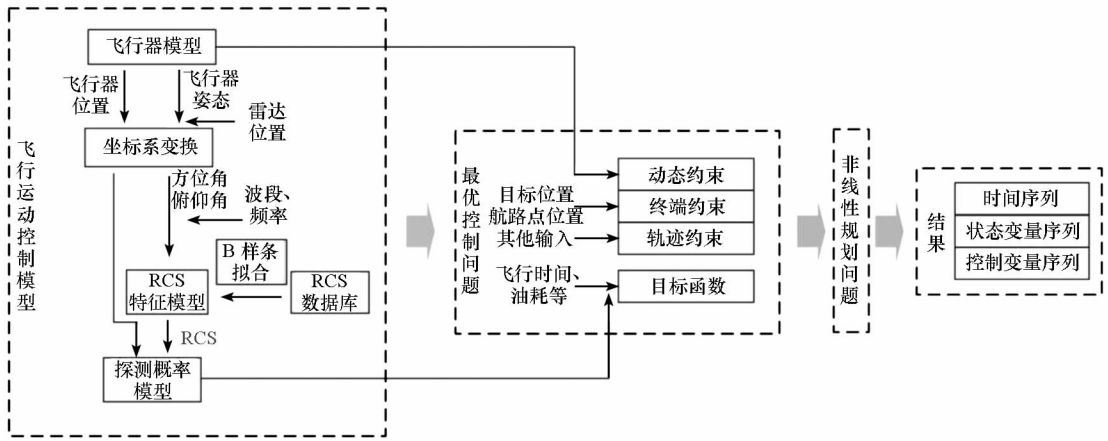


图 1 基于 RCS 的低可探测性轨迹优化问题求解框架

Fig. 1 Solving framework for Low Observable trajectory optimization problem based on RCS

$$J = w_t(t_f - t_0) + w_p \int_{t_0}^{t_f} PD(t) dt \quad (26)$$

其中,前一项表示飞行总时间,后一项表示探测概率的积分。 w_t 和 w_p 分别表示两者的权重系数。当满足 $w_p = 0$ 时问题转换为最小时间轨迹优化问题。然后转换为与式 (15) 相一致的形式为

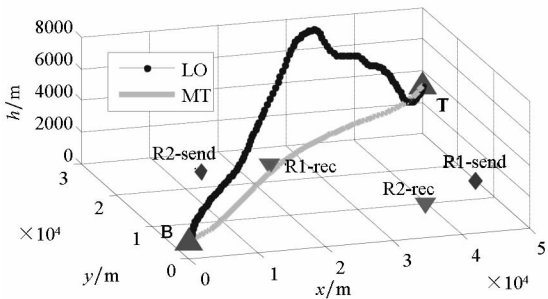
$$J = w_t(t_f - t_0) + \frac{(t_f - t_0)w_p}{2} \int_{-1}^1 PD(\tau) d\tau \quad (27)$$

4 仿真计算与结果分析

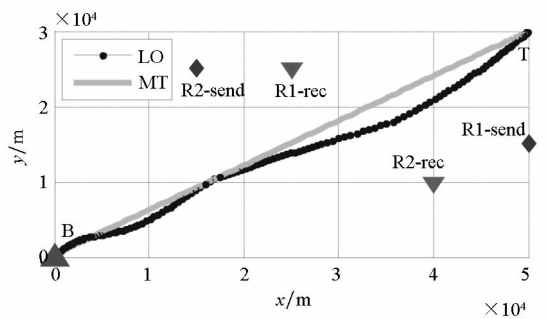
为了验证本文问题建模和求解方法的有效性,本节给出双基地雷达系统中低可探测性轨迹优化仿真实例,并与以最短飞行时间作为目标函数的轨迹优化结果进行比较分析。假设雷达的位置、类型、相关参数固定,探测概率 $PD \geq 0.1$ 时认为飞行器处于暴露状态。在仿真实例中,B、T 和 R 分别表示起始点、目标点和雷达所处位置。低可探测目标和最短飞行时间目标生成的两种轨迹分别表示为 LO(Low Observable) 轨迹和 MT(Minimum Time) 轨迹。为了便于比较 LO 轨迹和 MT 轨迹,实验结果仅给出了两者飞行控制量和状态量随 x 坐标的变化过程。由于飞行速度沿 x 方向的变化不大,此结果也可作为这些变量随时间变化结果的参考。

两组双基地雷达发射端分别位于 (50, 15, 0) km 和 (15, 25, 0) km,对应接收端位于 (25, 25, 0) km 和 (40, 10, 0) km。两个航路点位于 (0, 0, 1) km (起点) 和 (50, 30, 3) km (目标)。图 2 给出了优化生成的 LO 轨迹和 MT 轨迹,并且对相关变量随飞行器在 x 方向距离的变化关系进行了比较。由图 2(a)、(b) 可以看出规划出的生成飞行轨迹三维视图和俯视图。两条轨迹上飞行

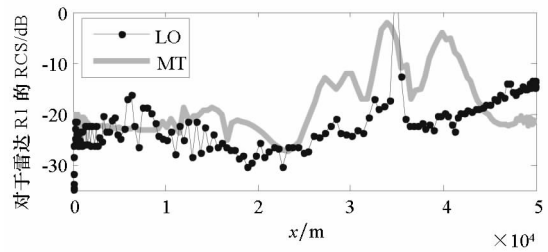
器 RCS 变化比较如图 2(c)、(d),可以看出 LO 轨迹上的 RCS 得到了普遍降低。表 1 给出了两条轨迹飞行时间和暴露时间对比结果,与 MT 轨迹相比,最大限度减少了飞行暴露时间。由图 2(e) 可以看出,LO 轨迹在大部分的时段保持在雷达探测概率 $PD \leq 0.01$ 。LO 轨迹姿态角和控制量变化分别如图 3(a) 和 3(b)。



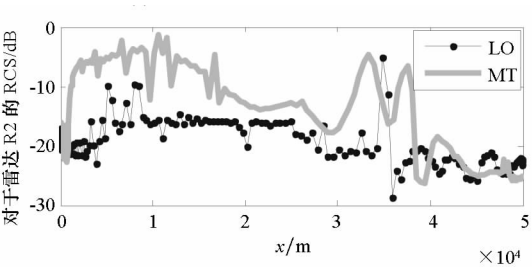
(a) 生成的飞行轨迹(三维视图)



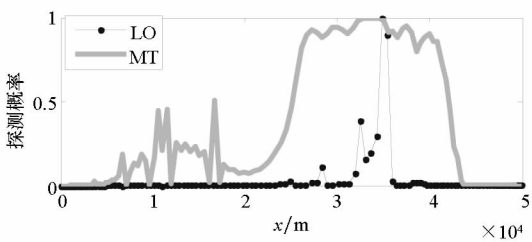
(b) 生成的飞行轨迹(俯视图)



(c) 相对于雷达 R1 的飞行器 RCS

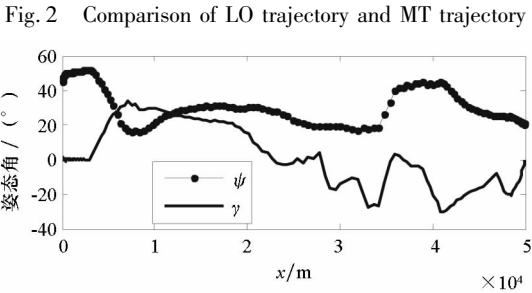


(d) 相对于雷达 R2 的飞行器 RCS

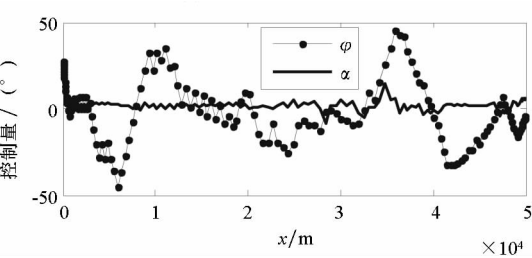


(e) 雷达探测概率

图2 LO 轨迹与 MT 轨迹比较



(a) LO 轨迹飞行姿态角



(b) LO 轨迹飞行控制量

图3 LO 轨迹求解结果

Fig. 3 The solving results of LO trajectory

表1 飞行时间和暴露时间比较

Tab.1 Comparison of flight time and exposure time

	暴露时间/s	飞行时间/s	暴露时间所占比例/%
MT 轨迹	98	167	58.7
LO 轨迹	22	231	9.5

5 结 论

本文主要建立了基于 RCS 和飞行姿态控制的三维低可探测性轨迹优化的求解框架。与传统的威胁环境下飞行器航迹规划相比,规划由平面

或三维航迹层深入到了带有状态量和控制量的三维轨迹层。将 Guass 伪谱方法应用于飞行器低可探测性轨迹优化问题,从而将连续的最优控制问题转换为非线性规划问题,实现对飞行器低可探测性轨迹优化问题的求解。通过仿真实例验证,在双基地雷达系统中,飞行轨迹能够保持较低的雷达探测概率,在实例中使飞行暴露时间占总飞行时间的比重小于 10%。

参考文献 (References)

[1] Inanc T. Framework for low-observable trajectory generation in presence of multiple radars[J]. Journal of Guidance, Control, and Dynamics, 2008, 31(6):1740-1749.

[2] Norsell M. Multistage trajectory optimization with radar range constraints[J]. AIAA Journal of Aircraft, 2005, 42(2): 849-857.

[3] Mark M. Real-time optimal trajectory generation for constrained dynamical systems [D]. Pasadena, California: California Institute of Technology, 2003.

[4] 丁晓东,刘毅,李为民. 基于动态 RCS 的无人机航迹实时规划方法研究[J]. 系统工程与电子技术, 2008, 30(5): 868-871.

DING Xiaodong, LIU Yi, LI Weimin. Dynamic RCS and real-time based analysis of method of UAV route planning[J]. Systems Engineering and Electronics, 2008,30(5): 868-871. (in Chinese)

[5] Kabamba P T, Meerkov S M, Zeitz F H. Optimal path planning for unmanned combat aerial vehicles to defeat radar tracking[J]. Journal of Guidance Control and Dynamics, 2006, 29(2):280-288.

[6] 苏抗,周建江. 有限姿控能力的低 RCS 微小卫星姿态实时规划[J]. 航空学报, 2010, 31(9): 1841-1848.

SU Kang, ZHOU Jianjiang. Real-time attitude planning for low RCS micro-satellites with limited attitude control ability[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(9): 1841-1848. (in Chinese)

[7] Benson D. A Gauss pseudospectral transcription for optimal control[D]. Cambridge: MIT, 2005.

[8] Philip E G. SNOPT: An SQP algorithm for large scale constrained optimization[J]. SIAM Review, 2005, 47(1): 99-131.

[9] Forsey D R, Bartels R H. Surface fitting with hierarchical splines[J]. ACM Transactions on Graphics 1995, 14(2): 134-161.

[10] 吴森堂,费玉华. 飞行控制系统[M]. 北京:北京航空航天大学出版社, 2005: 12-13.

WU Sentang, FEI Yuhua. Flight control system [M]. Beijing: Press of BUAA, 2005: 12-13. (in Chinese)

[11] Leonov S A, Leonov A I. Handbook of computer simulation in radio engineering, communications, and radar [M]. Norwood: Artech House, 2001: 158-160.

[12] 方学力,杨永祥. 雷达与雷达网的目标检测威力模型[J]. 现代雷达, 2008, 30(7): 18-20.

FANG Xueli, YANG Yongxiang. Radar target detection range model based on detection probability [J]. Modern Radar, 2008, 30(7): 18-20. (in Chinese)

空心 和 Halbach 永磁直线同步电机的牵引力和法向力分析*

刘恒坤, 张 晓, 弥 柱

(国防科技大学 机电工程与自动化学院, 湖南 长沙 410073)

摘 要: 针对由空芯线圈和 Halbach 永磁结构组成的直线同步牵引电机, 提出其牵引力和法向力的一种解析计算方法。首先从相同结构的旋转电机出发, 推导得到其牵引力和法向力的解析计算公式; 然后, 将所得结果推广到直线电机结构, 得到基于空芯线圈和 Halbach 永磁结构的直线同步电机牵引力和法向力的计算方法。所得解析解的有效性由有限元分析软件 Ansoft Maxwell 得到验证。此方法为基于空芯线圈和 Halbach 永磁结构的直线同步电机的结构设计和性能分析提供了基础。

关键词: 同步直线电机; Halbach 永磁体; 磁场; 有限元方法

中图分类号: TM351; TM359 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0094-04

Analytical computation of normal and tangential forces in linear synchronous motor with air-core and Halbach permanent magnets

LIU Hengkun, ZHANG Xiao, MI Zhu

(College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China)

Abstract: An analytical method of calculating the normal and tangential forces in linear synchronous motor with air-core and Halbach permanent magnets is developed. First, the analytical solution to the normal and tangential forces of the rotary motor with the same structure is deduced. Then, the proposed method is extended to linear motor, and the analytical method of computing the normal and tangential forces in linear synchronous motor with air-core and Halbach permanent magnets is obtained. Finally, the effectiveness of the proposed analytical method is validated by Ansoft Maxwell, which is a famous finite element analytical software. The proposed solution provides an important tool for structure design and performance analysis of linear synchronous motor with air-core and Halbach permanent magnets.

Key words: Linear Synchronous Motor; Halbach permanent magnets; magnetic field; finite element

永磁直线同步电机(PMLSM)因其动态性能好、可靠性高,越来越受到人们的关注^[1-3]。随着永磁直线同步电机的广泛应用,对其性能的要求也越来越高^[4]。永磁直线同步电机的牵引力波动是影响其性能的主要因素;此外,电机的法向力的大小也成为衡量电机性能的一个指标。对电机的力分析表明,若能得到标准的正弦气隙磁密,则能极大地减小电机的波动。采用常规磁体结构时不能得到理想的正弦波气隙磁密分布。为此,将 Halbach 磁体结构应用于永磁电机,这不仅可以使气隙磁密正弦化,而且气隙磁通也会增加,这意味着电磁牵引力增大^[5-7]。为进一步提高电机牵引力的平稳性,同时减弱电机法向力,初级采用空心线圈。常规带铁芯的同步直线电机,牵引力存在着6次谐波^[8-9],这会影响电机牵引力的平稳性。采用空心线圈作为初级,可以消除6次谐波的影响,使电机的牵引力更加平稳;另外,采用空心初级线圈,可以极大地减小初级与次级之间的

法向力,由此可以减弱电机对外设在法向力上的不利影响。

1 Halbach 结构永磁体介绍

1978年8月,美国物理学家 K. Halbach 提出了一种新颖的永磁体结构^[10-12],见图1。将这种结构用于电机便形成了一种特殊的永磁电机——Halbach 电机。这种结构电机的气隙磁密波形比常规磁体结构电机的气隙磁密波形具有更好的正弦性,并具有很好的转子轭部磁屏蔽作用,这对于提高电动机的转矩密度和功率密度极为有利^[13-14]。

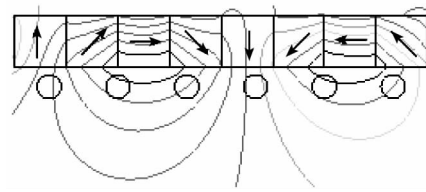


图1 Halbach 永磁结构图

Fig. 1 Halbach permanent magnet structure

* 收稿日期:2011-07-31

作者简介:刘恒坤(1975—),男,重庆江津人,讲师,博士,E-mail:liberry@sina.com

图1中,箭头为永磁的充磁方向,6个圆圈为三相线圈,封闭的曲线为磁力线。相邻两个永磁铁之间的充磁方向相差 45° 。由图可知,磁铁一侧磁力线分布较密,另外一侧几乎没有磁力线通过,这是 Halbach 结构磁场分布的重要特征。

2 电机的结构描述

本文研究的三相直线同步电机,其初级采用空心线圈,次极采用 Halbach 结构的永磁,采用软件 Ansoft Maxwell 建立电机模型如图2所示。[15]

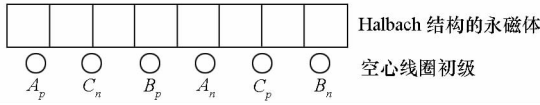


图2 有限元电机模型

Fig.2 Diagram of Finite element motor model

电机参数如下:极距 258mm,电机厚度 170mm,初级线圈采用铜线,其直径 30mm,永磁截面尺寸 64.5mm×64.5mm,永磁材料 NdFe48。

3 牵引力和法向力的计算

3.1 解析计算

直线电机可以由旋转电机剖开得到,为便于分析,这里以旋转电机为对象进行展开研究,不失一般性,带铁芯的旋转电机模型如图3所示。[2]

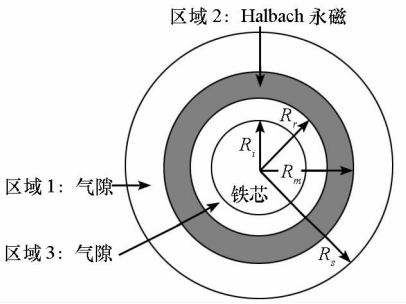


图3 Halbach 永磁电机结构示意图

Fig.3 Halbach permanent magnet motor

如图3中区域2是 Halbach 永磁,其余两个区域为空气。 R_s 、 R_m 、 R_r 和 R_i 分别为外铁芯半径、永磁外径、永磁内径和内铁芯半径。当 R_s 为无穷大,且 R_i 为0时,即为空心初级电机。计算时作如下假设:(1)磁体为理想 Halbach 结构;(2)忽略轴向长度的影响;(3)铁芯的磁导率无穷大。

在图3中建立极坐标系 $r-O-\theta$,其原点 O 位于电机的中心,径向坐标和圆周方向的坐标分别定义为 r 和 θ 。由于 Halbach 永磁结构磁化强度的分布 \mathbf{M} 呈正弦变化,其在极坐标系中可表

示为^[2]

$$\mathbf{M} = \frac{B_r}{\mu_0} \cos p \theta \mathbf{e}_r \pm \frac{B_r}{\mu_0} \sin p \theta \mathbf{e}_\theta \quad (1)$$

其中 p 为电机的极对数,负号代表永磁磁场在永磁体的外部,正号代表永磁磁场在永磁体的内部, \mathbf{e}_r 和 \mathbf{e}_θ 分别是沿坐标系径向和圆周方向的单位矢量。

图3中, Halbach 永磁结构在区域1、2和3中产生的磁场可由标量磁势进行表示。

区域1内, $R_m < r < R_s$, 磁势 φ_1 满足^[2]:

$$\nabla^2 \varphi_1 = \frac{\partial^2 \varphi_1}{\partial r^2} + \frac{1}{r} \frac{\partial \varphi_1}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \varphi_1}{\partial \theta^2} = 0 \quad (2)$$

区域2内, $R_r < r < R_m$, 磁势 φ_2 满足^[2]:

$$\nabla^2 \varphi_2 = \frac{\partial^2 \varphi_2}{\partial r^2} + \frac{1}{r} \frac{\partial \varphi_2}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \varphi_2}{\partial \theta^2} = \frac{\nabla \mathbf{M}}{\mu_r} \quad (3)$$

区域3内, $R_i < r < R_r$, 磁势 φ_3 满足^[2]:

$$\nabla^2 \varphi_3 = \frac{\partial^2 \varphi_3}{\partial r^2} + \frac{1}{r} \frac{\partial \varphi_3}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \varphi_3}{\partial \theta^2} = 0 \quad (4)$$

式(2)~(4)的通解分别为

$$\varphi_1(r, \theta) = (C_1 r^p + C_2 r^{-p}) \cos p \theta \quad (5)$$

$$\varphi_2(r, \theta) = \left(C_3 r^p + C_4 r^{-p} + \frac{B_r}{\mu_0 \mu_r} \frac{r}{1 \pm p} \right) \cos p \theta \quad (6)$$

$$\varphi_3(r, \theta) = (C_5 r^p + C_6 r^{-p}) \cos p \theta \quad (7)$$

图3的电机结构具有如下边界条件:

$$H_{\theta 1} = 0, \quad (r = R_s) \quad (8)$$

$$B_{r1} = B_{r2}, \quad (r = R_m) \quad (9)$$

$$H_{\theta 1} = H_{\theta 2}, \quad (r = R_m) \quad (10)$$

$$B_{r2} = B_{r3}, \quad (r = R_r) \quad (11)$$

$$H_{\theta 2} = H_{\theta 3}, \quad (r = R_r) \quad (12)$$

$$H_{\theta 3} = 0, \quad (r = R_i) \quad (13)$$

磁势 φ 和磁通 H 满足以下条件:

$$H_r = -\frac{\partial \varphi}{\partial r}, \quad H_\theta = -\frac{1}{r} \frac{\partial \varphi}{\partial \theta} \quad (14)$$

结合式(1)~(14)即可求出各个区域的磁场强度和磁通密度。观察图2和图3可知,图2中直线电机初级线圈所处的磁场环境即为图3中旋转电机的区域1,对应到本文所指的空心 Halbach 直线电机模型,即 $R_s \rightarrow \infty$, $R_i = 0$, 由于篇幅原因,这里只给出区域1内磁通密度表达式:

$$B_{r1} = \mu_0 H_{r1} = -\mu_0 \frac{\partial \varphi_1}{\partial r} = B_0 \cos p \theta \quad (15)$$

$$B_{\theta 1} = \mu_0 H_{\theta 1} = -\mu_0 \frac{1}{r} \frac{\partial \varphi_1}{\partial \theta} = B_0 \sin p \theta \quad (16)$$

其中:

$$B_0 = \frac{-\frac{2B_r p}{1+p}(1+\mu_r)\left[1-\left(\frac{R_r}{R_m}\right)^{p+1}\right]\left(\frac{R_m}{r}\right)^{p+1}}{(1-\mu_r)^2\left(\frac{R_r}{R_m}\right)^{2p}-(1+\mu_r)^2} \quad (17)$$

图 2 中的三相电流可分别表示为

$$I_a = A \sin(2\pi f \cdot t + \theta_0) \quad (18)$$

$$I_b = A \sin(2\pi f \cdot t + \theta_0 - \frac{2\pi}{3}) \quad (19)$$

$$I_c = A \sin(2\pi f \cdot t + \theta_0 + \frac{2\pi}{3}) \quad (20)$$

其中 A 为电流的幅值, f 是电流的变化频率, θ_0 为初始相位。

在图 2 中, 定义 X 轴水平向右, Y 轴垂直向上, 则各线圈沿 X 轴的受力表现为电机的牵引力, 沿 Y 轴的受力表现为电机的法向力。考虑 Halbach 磁场表达式 (15) 和 (16), 利用通电线圈在磁场中的受力公式, 直线电机中各线圈受的牵引力和法向力可分别计算为

$$F_{apx} = B_0 \cos \theta_{ap} \cdot I_a \cdot L \quad (21)$$

$$F_{apy} = B_0 \sin \theta_{ap} \cdot I_a \cdot L \quad (22)$$

$$F_{anx} = -B_0 \cos \theta_{an} \cdot I_a \cdot L \quad (23)$$

$$F_{any} = -B_0 \sin \theta_{an} \cdot I_a \cdot L \quad (24)$$

$$F_{bpx} = B_0 \cos \theta_{bp} \cdot I_b \cdot L \quad (25)$$

$$F_{bpy} = B_0 \sin \theta_{bp} \cdot I_b \cdot L \quad (26)$$

$$F_{bnx} = -B_0 \cos \theta_{bn} \cdot I_b \cdot L \quad (27)$$

$$F_{bny} = -B_0 \sin \theta_{bn} \cdot I_b \cdot L \quad (28)$$

$$F_{cpx} = B_0 \cos \theta_{cp} \cdot I_c \cdot L \quad (29)$$

$$F_{cpy} = B_0 \sin \theta_{cp} \cdot I_c \cdot L \quad (30)$$

$$F_{cnx} = -B_0 \cos \theta_{cn} \cdot I_c \cdot L \quad (31)$$

$$F_{cny} = -B_0 \sin \theta_{cn} \cdot I_c \cdot L \quad (32)$$

其中 L 为电机的宽度。各式中 θ_0 的变化可以改变电机的功角。由于力的相互作用, 永磁体所受的力的大小即为空心线圈所受的力, 只不过方向相反, 这里不考虑力的方向, 根据式 (21) ~ (32) 可知, 图 2 所示结构的直线同步电机产生的牵引力和法向力可计算为

$$F_x = F_{apx} + F_{anx} + F_{bpx} + F_{bnx} + F_{cpx} + F_{cnx} \quad (33)$$

$$F_y = F_{apy} + F_{any} + F_{bpy} + F_{bny} + F_{cpy} + F_{cny} \quad (34)$$

图 2 所示的直线电机结构可以看作是由图 3 所示的旋转电机展开得到的, 式 (1) ~ (34) 的推导可推广应用到图 2 的直线电机结构。

下面给出一个计算实例。选取图 2 所示电机结构中, Halbach 永磁体的高度为 64.5mm, 单周期阵列包括 8 块永磁体, 长度为 516mm。从而等效到图 3 所示的电机结构中, 有

$$R_r = \frac{p \times 0.516}{2\pi}$$

$$R_m = R_r + 0.0645$$

$$r = R_m + 0.025$$

另外选取永磁材料的参数为

$$B_r = 1.2603, \quad \mu_r = 1.085$$

线圈中电流的参数为

$$A = 1414, \quad f = 50$$

电机的宽度 $L = 170\text{mm}$ 。

当电机位于图 2 所示的位置时

$$\theta_{ap} = 0, \quad \theta_{cn} = \frac{\pi}{3}, \quad \theta_{bp} = \frac{2\pi}{3},$$

$$\theta_{an} = \pi, \quad \theta_{cp} = \frac{4\pi}{3}, \quad \theta_{bn} = \frac{5\pi}{3}.$$

图 4 为不同电机极对数 p , 功角 θ 在 $0 \sim 360^\circ$ 范围内变化时, 一个周期 Halbach 永磁体所受力的仿真结果。当电机极对数 p 较少时, 容易理解, 图 2 所示的直线电机与图 3 所示旋转电机的误差较大。为此, 需要取合适的电机极对数 p , 以减小计算误差。

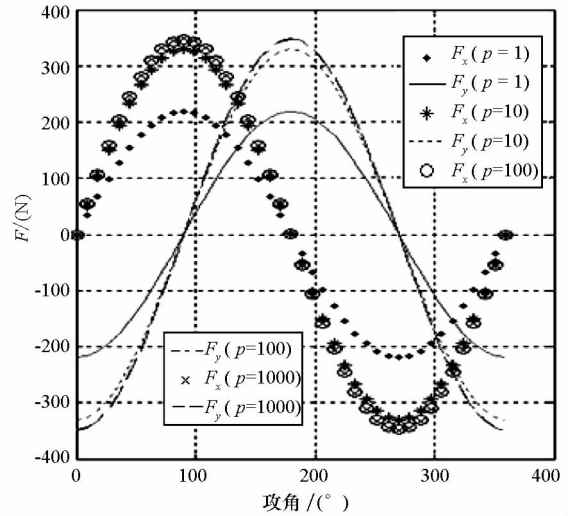


图 4 极对数 p 变化时的结果

Fig. 4 Diagram of result while p is varying

图 4 中, F_x 为牵引力, F_y 为法向力。由图可知, 牵引力为正弦分布, 法向力为余弦分布。从图中可以看出, 当电机极对数 p 较大时, 如 p 为 100 以上时, 牵引力和法向力随极对数增加变化缓慢。

3.2 有限元结果

由于初级和次级沿垂直于次级运动方向的每一横截面的形状均相同, 因此采用 2D 长定子模型进行分析。利用 Ansoft Maxwell 软件, 取一个周期的 Halbach 永磁, 建立 2D 模型。初级与次级的相对位置如图 2 所示, 功角 θ 在 $0 \sim 360^\circ$ 范围内变化时, 求出永磁体的受力有限元结果, 并将 p 取

1000 时的解析结果与之对比,二者的对比结果见图 5。

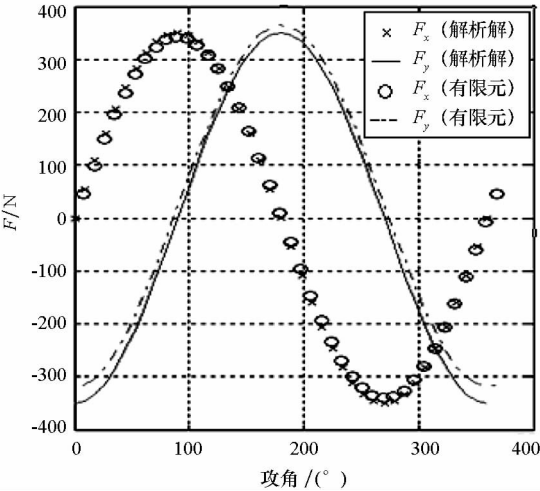


图 5 有限元与解析解的对比

Fig. 5 Finite element result and analytical result

图 5 中, F_x 表示牵引力, F_y 表示法向力,由有限元与解析解比较可知,二者所得到的永磁体的受力情况基本一致。两者之间误差的原因在于,解析计算时认为永磁磁场是理想的正弦磁场,同时忽略了线圈切割永磁磁场产生的电势及直线电机端部效应的影响。

4 结 论

对于空心 Halbach 永磁直线电机,给出了其模型。利用旋转电机模型进行解析分析,得到气隙磁场分布的解析解,进而得出直线电机磁场分布的解析结果,计算出永磁体的受力情况。利用有限元计算结果与解析结果进行对比,验证了二者的一致性。解析计算方法相对于有限元方法而言,计算量小,计算时间少,设计方便,这种方法在永磁直线同步电机设计的初期具有很强的实用性;在设计电机时,根据电机的性能指标,利用这种方法,能够快速计算出电机的参数,然后再利用有限元方法对电机参数进一步优化,提高了电机设计的效率。

参考文献 (References)

[1] 毛军红, 罗俊航, 姜强, 等. 一种永磁直线电机的永磁体阵列设计[J]. 西安交通大学学报, 2007, 41(3): 353 – 357.
MAO Junhong, LUO Junhang, JIANG Qiang, et al. Design of magnet arrays for a permanent magnetic linear motor[J]. Journal of Xi'an Jiao Tong University, 2007, 41(3): 353 – 357. (in Chinese)

[2] 何仁, 牛润新, 董颖. 永磁磁轨制动技术在轨道交通中的应用[J]. 中国安全科学学报, 2007, 17(1): 167 – 171.
HE Ren, NIU Runxin, DONG Yin. Application of permanent magnet track braking technique in railway[J]. China Safety Science Journal, 2007, 17(1): 167 – 171. (in Chinese)

[3] 宋国强, 赵朝会. Halbach 列永磁电机的研究[J]. 电机与控制应用, 2008, 35(7): 12 – 17.
SONG Guoqiang, ZHAO Chaohui. Study of Halbach array PM machines[J]. EMCA, 2008, 35(7): 12 – 17. (in Chinese)

[4] 崔皆凡, 蒋莉莉, 王贺敏, 等. Halbach 磁体结构应用于永磁直线同步电机的研究[J]. 沈阳工业大学学报, 2007, 29(4): 400 – 403.
CUI Jiefan, JIANG Lili, WANG Hemin, et al. Research on permanent magnet linear synchronous motor with Halbach array[J]. Journal of Shenyang University of Technology, 2007, 29(4): 400 – 403. (in Chinese)

[5] Xia Z P, Zhu Z Q, Howe D. Analytical magnetic field analysis of Halbach magnetized permanent magnet machines[J]. IEEE Transactions on Magnetics, 2004, 40(4): 1864 – 1872.

[6] Moon G L, Sung Q L, Gweon D G. Analysis of Halbach magnet array and its application to linear motor [J]. Mechatronics, 2004, 14(5): 115 – 128.

[7] 潘孟春, 陈棣湘, 罗飞路, 等. 常导高速磁悬浮列车电磁场的分析与测量[J]. 铁道学报, 2004, 26(4): 126 – 128.
PAN Mengchun, CHEN Dixiang, LUO Feilu, et al. Electromagnetic field analysis and measurement of high speed normal magnetic levitation vehicle system[J]. Journal of the China Railway Society, 2004, 26(4): 126 – 128. (in Chinese)

[8] 陈棣湘, 潘孟春, 罗飞路, 等. 高速磁悬浮列车电磁场的模拟计算[J]. 国防科技大学学报, 2003, 25(4): 76 – 78.
CHEN Dixiang, PAN Mengchun, LUO Feilu, et al. The simulated calculation on electromagnetic fields of the high speed magnetic levitation vehicle[J]. Journal of National University of Defense Technology, 2003, 25(4): 76 – 78. (in Chinese)

[9] 徐衍亮, 姚福安, 房建成. Halbach 磁体结构电动机及其常规磁体结构电动机比较研究(Ⅱ)[J]. 电工技术学报, 2004, 19(6): 58 – 62.
XU Yanliang, YAO Fuan, FANG Jiancheng. Halbach array permanent magnet machine and its performance comparison with the normal array one(Ⅱ)[J]. Transactions of the China Electrotechnical Society, 2004, 19(6): 58 – 62. (in Chinese)

[10] Merritt B T, Post R F, Reifuerst G R, et al. Halbach arraymotor/generators-A novel generalized electric machine [C]// Halbach Festschrift Symposium, 1995.

[11] Halbach K. Physical and optical properties of rare earth cobalt magnets [J]. Nuclear Instruments and Methods, 1981, 187: 109 – 117.

[12] Halbach K. Perturbation effect in segmented rare earth cobalt magnets [J]. Nuclear Instruments and Methods, 1982, 198: 213 – 215.

[13] Zhu H Q, Xia Z P, Atallah K, et al. Powder alignment system for anisotropic bonded NdFeB Halbach cylinders [R]. 2000 IEEE, 0018 – 9464.

[14] Zhu Z Q, Howe D. Halbach permanent magnet machines and applications2a review [C]// Proceedings Inst. Elect. Eng. Electric Power Applications, 2001, 148(4): 299 – 308.

[15] 刘国强, 赵凌志, 蒋继姬. Ansoft 工程电磁场有限元分析[M]. 北京: 电子工业出版社, 2006.
LIU Guoqiang, ZHAO Lingzhi, JIANG Jiya. Ansoft engineering electromagnetic field analysis [M]. Beijing: Electronic Engineering Press, 2006. (in Chinese)

面向 SDR 应用的向量存储器的设计与优化*

陈海燕,刘 胜,刘 仲,陈书明
(国防科技大学 计算机学院,湖南 长沙 410073)

摘 要: 针对面向 SDR 应用的 SIMD 数字信号处理器高带宽数据访存需求,提出并实现了一种新型的向量存储结构。该向量存储器由 16 路向量存储块构成,每路采用两组多体低位地址交叉编址存储结构,减少了访存体冲突,充分利用多存储体带宽,以较小的功耗代价实现并行访问多个向量数据。在此基础上,还设计了一种向量访存重整单元,使向量存储器可灵活支持多路 SIMD 结构向量处理单元的非对齐访问,实现了其对向量存储器的共享。测试结果表明,该向量存储器能有效减少或消除向量处理单元之间的数据混洗操作,加速相关应用算法。

关键词: 向量处理;单指令流多数据流;访存冲突;多存储体交叉;混洗;非对齐访问
中图分类号: TP368.1 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0098-05

Design and optimization of the vector memory applying for SDR

CHEN Haiyan, LIU Sheng, LIU Zhong, CHEN Shuming
(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: To meet the high memory bandwidth of SDR-oriented SIMD DSPs, a novel Vector Memory (VM) architecture is proposed. The VM consists of 16-way Vector Memory Blocks (VBs), and each VB contains two groups of multi-bank memory structure with low-order interleaved addressing. This structure aims at reducing the memory access conflicts, making best use of the bandwidth of the multi-bank memory, and realizing the parallel vector data access at the cost of low power consumption. Besides, a vector rearrangement unit is designed and implemented in the VM to support the 16-way unaligned SIMD vector access and share the VM space. Experimental results show that the proposed VM architecture can efficiently reduce or eliminate the data shuffling operations and speed up SDR applications.

Key words: vector processing; SIMD; memory access conflict; multi-bank interleave; shuffle; unaligned access

随着无线通信技术的不断发展,通信协议的快速演进以及多种通信协议间的互操作性要求,传统的基于硬件的 ASIC 加速器的解决方案代价高昂而且复杂,越来越不能满足应用需求^[1],软件无线电 (SDR, Software Defined Radio) 技术成为业界新的应用研究热点。由于 DSP 强大的计算能力、较低的功耗预算和灵活的可编程性,可更好支持未来软件升级、降低硬件成本,将其作为核心处理引擎的 SDR 技术平台成为支持无线通信多模应用的理想解决方案。

SDR 应用面向音频、视频等流媒体数字信号处理,是典型的高计算密集型应用,具有很高的数据并行性。为了满足其应用需求,越来越多的高性能 DSP 采用单指令流多数据流 (SIMD) 结构,单片集成多宽度向量处理单元来开发应用的数据级并行,以提供强大的实时数据处理能力。多宽度 SIMD 结构处理器中的向量处理单元需要更高

的访存带宽,面临着更为严重的“存储墙”问题^[2]。

由于流媒体数据的时空局部性较差,通常的 Cache 存储层次结构不能很好匹配其访存特点;而且 Cache 存储层次结构中的访存缺失造成的流水线停顿将严重降低多宽度 SIMD 结构 DSP 的处理性能,无法满足实时处理要求。因此 SIMD 结构 DSP 一般配置了大容量片内存储器,尽量将运算数据保存在片内存储器中,以避免频繁的外存访问^[3],降低访存延迟。如何为向量处理单元提供持续、充足的数据访存带宽,提高算法的访存效率和降低功耗,成为存储系统设计面临的重要问题。为了高效完成 SIMD 操作,数据必须按向量处理单元对齐,而有的存储系统仅提供按 SIMD 宽度地址对齐的访问,如 Stanford 的流处理器 Imagine^[4],每个运算处理单元 (PE, Process Element) 只能访问自己对应的存储体,对于较复

* 收稿日期:2011-06-02
基金项目:国家“核高基”重大专项 (2009ZX01034-001-001-006)
作者简介:陈海燕 (1967—),女,四川南充人,研究员,硕士,硕士生导师,E-mail:hychen608@163.com

杂的数据访存模式,PE 间的数据交互需要大量的混洗、打包、解包等额外操作^[5];有的在存储器和 PE 间提供全交叉开关网络,如 AnySP^[6],但随着 SIMD 宽度的增加,交叉开关的时延和功耗也大幅增长。

本文针对面向 SDR 应用的一款多宽度 SIMD 结构的高性能 DSP 设计了一种新型的向量存储器,该向量存储器由 16 路向量存储块构成,每路包含两组采用多体低位交叉编址的 4 个单口 SRAM 存储体,不仅有利于降低功耗,还能减少访存冲突次数,支持多路并行向量访存。在此基础上,还实现了一种向量访存重整理单元,不仅支持常规地址对齐的向量数据访存,还以较小的硬件代价实现了非对齐方式的向量访问,可大幅减少或完全消除 SDR 应用中一些常见算法的混洗操作,因而压缩了代码密度,提高了向量数据访存效率。

1 相关工作

1.1 YHFT-Matrix 结构框图

根据 SDR 多模应用需求,我们自主研发了一款高性能 16 路 SIMD 结构的 32 位数字信号处理器 YHFT-Matrix,设计主频 500MHz,8/16/32 位定点峰值性能为 32/16/8GMAC/s。YHFT-Matrix 采用标量处理单元(SPU)和向量处理单元(VPU)并行结构,由共同的取值派发部件完成标、向量指令派发,该 DSP 采用 10 发射超常指令字(VLIW)执行包结构,包含 5 条标量和 5 条向量指令。SPU 负责标量指令的执行,完成标量数据处理;VPU 负责向量指令的执行,完成向量数据运算。为同时满足指令、标量和向量数据并行访问,片上存储器分为标量存储器及向量存储器。标量存储器实现指令访问和标量数据访问,由标、向量处理单元共享的指令 Cache(ICache)和标量数据 Cache(DCache)组成;向量存储器主要实现 VPU 的向量数据访问,由 16 路向量存储块构成。VPU 由 16 个同构的 PE 阵列构成,每个 PE 集成了 MAC、ALU 和位处理 3 个子功能部件,分别对应执行 VLIW 指令包中的一条向量指令;其余两条向量指令为向量访存指令,由向量存储器(VM)完成;所有 PE 按 SIMD 的方式执行向量指令。片内存储器通过 DMA 与外设或外部存储器进行数据交换。该 DSP 的总体结构框图如图 1 所示。图中,VM 除了可支持两条向量访存指令,还可支持 SPU 及 DMA 的并行访问。

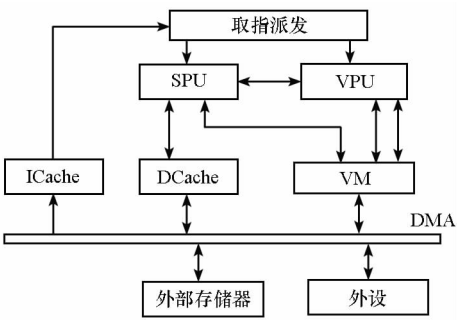


图 1 YHFT-Matrix 结构框图
Fig. 1 Framework of YHFT-Matrix DSP

1.2 向量存储器 VM 的总体结构

为了满足 VPU 内 16 个 PE 中 MAC 部件的数据带宽要求,VM 需要并行支持两条向量访存指令操作(LS0、LS1),同时还支持标量处理单元 SPU、DMA 对向量数据的访问,即 VM 需要支持 4 请求并行访问。VM 由专用的向量地址产生单元(VAGU)、与 PE 总数相同的向量存储块(VB0 ~ VB15)、向量访存控制器和向量输出选择器等主要功能模块组成,其组成结构及接口框图如图 2 所示。VAGU 实现向量访存指令的译码和地址计算;向量访存控制器实现 4 请求访存流水线的控制,即对仲裁、访存、输出选择和写回等各访存流水线站进行控制;向量输出选择器将读出的向量数据输出至各请求端。

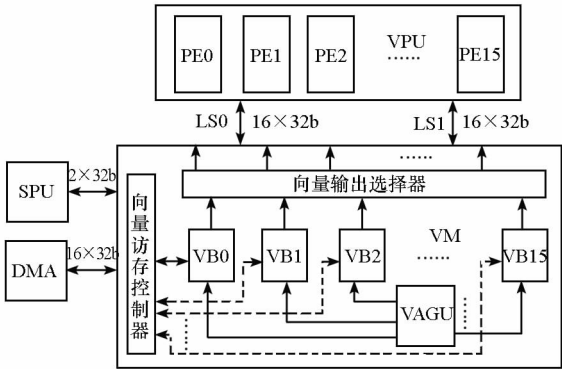


图 2 VM 的组成结构
Fig. 2 Components of VM

VM 总容量设计为 1MB,16 个 VB 完全同构,单个 VB 容量为 64KB,访存地址按 16 个 VB 低位地址交叉编址。由于嵌入式应用对 DSP 功耗的严格限制,每个 VB 均选择单端口 SRAM 作为存储单元。为支持 4 路并行访问,减少并行访存冲突,每个 VB 由 8 个单端口 SRAM(每个大小为 8KB)存储体构成,按两组 4 体低位交叉编址组织成上下两个存储空间,上半存储区由 bank0 ~ bank3 构成,下半存储区由 bank0' ~ bank3' 构成。理论上,访存地址不存在体冲突时,每个 VB 最多

可同时响应 8 个访存请求。假设某个 VB 的访存地址为 ADDR, 如果 $(\text{ADDR}) \bmod (16 \times 4)$ 的值不同, 就可实现无冲突的并行访存; 如果将 DMA 访问和内核访问分布在上下不同的存储区, DMA 与内核则不存在访存冲突, 而且如果待处理数据的 DMA 传输和向量数据访存并行, 则可隐藏大部分 DMA 访问延时。

2 集中式向量地址产生单元

为了支持多种寻址模式和有符号/无符号的字节、半字、字等多粒度的向量数据访问, 专门的地址产生单元的设计非常重要。对于 16 路 SIMD 结构的 DSP, 我们设计了一个高带宽的向量地址产生单元 (VAGU), 实现两条向量访存指令的译码和地址计算功能。VAGU 根据支持的寻址模式, 采用集中式方式进行向量访存地址的计算, 再根据向量访存地址按访问粒度生成 16 个 VB 的访问地址和其他访存数据。

3 向量访存控制器的设计

访存控制器控制访存流水线的正确执行, 决定着流水线的性能。对于片上大容量多路向量存储器, 既要支持多请求并行访存流水, 又要保证性能, 向量访存控制器采用了分级的控制方法, 设计总控模块和 16 个 VB 的访存控制器 VMC0 ~ VMC15。总控模块负责 16 个 VB 的访存同步, 即访问请求仲裁开始的同步及向量数据写回的同步; VMC0 ~ VMC15 分别控制 16 个 VB 的访存流水线, 实现对单个 VB 的访问控制。

整个向量访存流水线分为 6 站: 地址计算站、仲裁站、访存译码站、访存站、输出选择站和写回站。其中, VAGU 在地址计算站完成地址计算和访存请求数据生成; 仲裁站负责接收和仲裁 VPU、SPU 和 DMA 对 VM 的访存请求, 并在发生访存地址冲突时缓存访存请求、选择访存信号并输出到下一站; 访存译码站负责将访存信号译码到 VB 内部各存储体端口, 准备下一拍的访存操作。对于 VPU 的向量读访存请求, 还需要输出选择站对各 VB 读出的数据进行选择和对齐整理, 并在写回站输出至各请求端。

多请求并行访问的关键是要解决访存冲突问题, 在 VM 中这是由向量访存控制器在仲裁站完成的。访存冲突主要发生在 VB 的 4 个请求的访问地址产生存储体冲突的情况。YHFT-Matrix 为 DMA 请求设置了 2 位可编程优先级, 当 DMA 和其他访存请求发生冲突时, 若 DMA 为高优先级

时 (优先级 > 0), 则先处理 DMA 请求, 并暂停其他内核请求; 否则先处理内核请求, 并暂停 DMA 请求。由于内核访问需要同步, 同时为了简化仲裁逻辑, 仲裁控制器中设计了两个状态机: 内核向量访问 (LS0、LS1、SPU) 请求仲裁状态机和内核向量访问与 DMA 访问的仲裁状态机, 其状态转换图分别见图 3、图 4。

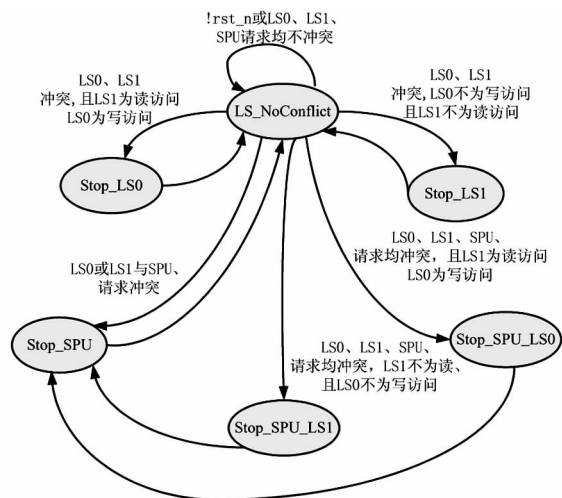


图 3 内核向量访存仲裁状态机

Fig. 3 State transition diagram for VB access arbitration of DSP core

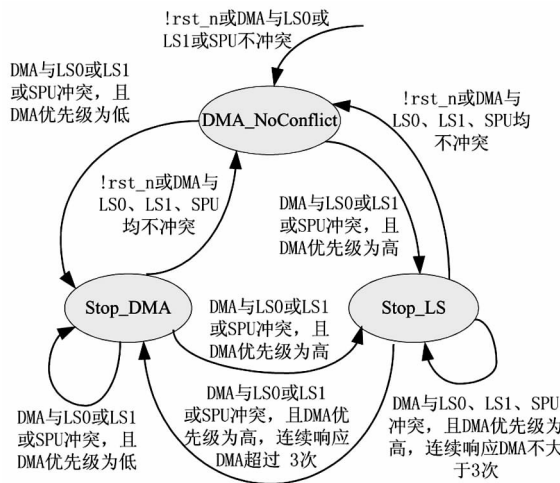


图 4 DMA 与内核向量访存仲裁状态机

Fig. 4 State transition diagram for VB access arbitration between DSP core and DMA

在仲裁 3 个内核访问时, YHFT-Matrix 设置了以下仲裁原则: 当向量 LS0/LS1 与 SPU 请求发生访存地址冲突时, 将按照向量 LS0/LS1 为高, SPU 为低的优先级进行串行访问; 当两个向量访问 LS0/LS1 发生访存冲突时, 则读访问优先, 若均为读或写访问, 则 LS0 具有较高优先级。因此, 图 3 所示的实现 VB 的 3 个内核访问请求仲裁的状态机共有 6 个状态: 内核访问无冲突状态 LS_NoConflict、SPU 访问停顿状态 Stop_SPU、向量访

问 LS0 停顿状态 Stop_LS0、向量访问 LS1 停顿状态 Stop_LS1、SPU 和 LS0 同时停顿状态 Stop_SPU_LS0、SPU 和 LS1 同时停顿状态 Stop_SPU_LS1。如果仅存在内核访存冲突,则访问会延迟 1~2 拍完成。

图 4 所示状态机只判断 DMA 请求是否与 3 个内核访问冲突,用 3 个状态表示:即不冲突则进入 DMA_NoConflict 状态,DMA 请求与内核访问可并行;存在冲突时,根据 DMA 优先级,进入 DMA 请求停顿状态 Stop_DMA,或内核访问停顿状态 Stop_LS,此时为避免内核请求得不到响应,设计规定当 DMA 优先级为高时,可连续响应 DMA 请求的次数由其优先级确定,即按其最高优先级算,最多可连续响应 3 次;若 DMA 优先级被设置为 0,则内核访存始终具有高优先级。

为了减少 DMA 和 VPU 的访存冲突,DMA 和 VPU 处理的数据应该尽量放在一个 VB 中上下两个不同的存储区,这样可以使外部数据加载和内核计算并行,减少访存冲突,隐藏外部数据加载延时,提高向量数据处理的性能。

4 支持非对齐向量访问的优化设计

为了高效完成 SIMD 操作,一般 N 路的 SIMD 结构处理器要求向量访存数据按 N 进行边界对齐,即向量存储器的各个向量存储块 VB 一般与相应的运算处理单元 PE 一一对应。对于访存模式复杂的应用,各 PE 之间需要的数据交互则完全由混洗等指令完成,使操作数的组织过程复杂化,严重降低了实际的向量访存效率^[5]。因此折中考虑性能和硬件实现代价,本文提出了一种优化的 VM 结构,支持向量数据的非对齐连续访问,使一个 PE 能有条件地访问到 VM 内的所有 VB,实现所有 PE 对 VM 存储空间共享,大幅提高相关应用算法的访存效率。

在这种优化的 VM 结构中,设计了一种新的向量访存重整单元,即在向量地址产生单元 VAGU 中增加了实现地址对齐整理功能的向量地址整理单元 (VARU, Vector Address Reorder Unit),在向量输出选择器中增加了数据输出对齐整理功能的向量数据整理单元 (VDRU, Vector Data Reorder Unit)。在地址计算站,VARU 根据 VAGU 计算出的向量地址,完成向量地址分解、复制和重整,即对向量访存地址进行判断,对于地址未对齐的访存请求进行循环移位(按 PE0~PE15 的顺序)和 VB 地址跨行加 1 操作,使得向量访存操作可以从任意非 N 路对齐的位置开始 N

个数据的连续访问,同时将该移位信息传递到下面的访存流水线。若本次操作为向量读操作,则在 VM 的输出选择站,需要 VDRU 根据流水线传递的移位信息对读出数据进行与 VARU 相反的循环移位操作后,再输出到对应的 PE 中。增加的 VARU 和 VDRU 这两个功能单元都主要实现循环移位功能,均由一个 16 项输入的循环移位器实现,硬件实现代价小,并未降低对应访存流水线的性能。

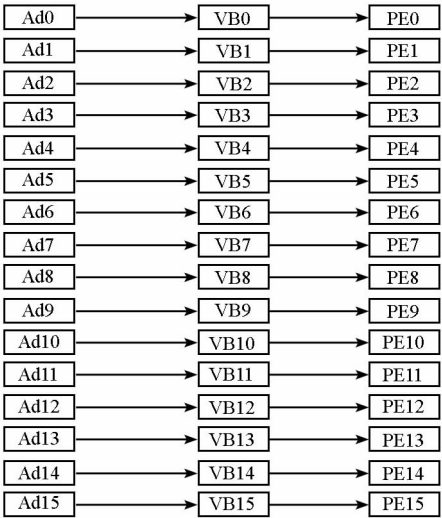


图 5 向量访存地址按 16 路对齐时 PE 对 VB 的读访问
Fig. 5 16-way aligned address load for PEs

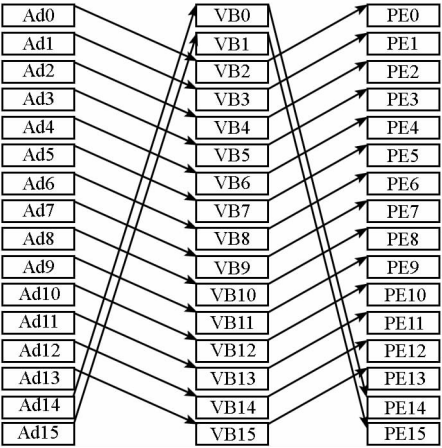


图 6 向量访存地址 = 2 时 PEs 对 VBs 的非对齐读访问
Fig. 6 Unaligned address load for PEs

图 5 和图 6 分别为向量读访存地址按 16 路边界对齐和向量读访存地址等于 2 时的向量读操作中 16 个 VB 与 16 个 PE 的关系流图,其中 Ad0~Ad15 为 VAGU 产生的 16 个访问 VB 的地址。可见,通过向量访存重整单元,VM 可支持地址对齐和非对齐方式的访问,每个 PE 可以访问 VM 的任何一个 VB。

下面以 SDR 中常用的有限滤波算法 (FIR) 为例,分析这种非对齐访问的访存性能。

基本的 FIR 算法如下：

$$y(n) = \sum_{i=0}^{M-1} h(i)x(n-i)$$

假设样本数 $n=1024$, 抽头 $M=16$, 数据均为定点 32 位, YHFT-Matrix 采用 16 路 SIMD 操作方式, 计算出全部 $y(n)$, 系数 $h(i)$ 需要 1 条向量 load 指令, 样本数 $x(n-i)$ 加载需要 1024 条向量 load 指令, 保存结果需要 1024 条向量 store 指令, 即可完成算法的访存操作。如果 VM 只支持 SIMD 宽度对齐的向量数据访问, 样本数 $x(n-i)$ 的加载只需要 $1024/16=64$ 条向量 load 指令, 比前者减少了 960 条, 但会额外增加 15 种混洗模式共 960 条混洗指令, 而每次混洗操作都需要使用向量 load 指令配置混洗模式, 所以两者向量 load/store 指令条数相同, 但前者无需混洗操作, 后者不仅会增加 960 条混洗指令, 而且多 SIMD 宽度 PE 间的混洗操作开销较大, 其软件流水实现困难, 无疑将大幅降低系统的计算效率。

5 性能分析与结论

设计使用 Synopsis 公司的 Design Compiler, 基于某厂家 65nm 工艺库进行逻辑综合, 结果表明, VM 的工作主频可达 500MHz 以上, 因此 VM 最大可同时提供 512Gbps 的向量、32Gbps 的标量、256Gbps 的 DMA 访存带宽。

目前基于 4 个 YHFT-Matrix 的多核 DSP 芯片已投片成功。下面在该 DSP 芯片的测试环境下, 基于其单核 YHFT-Matrix 的向量运算结构, 使用 SDR 中音频、视频信号处理中常用的有限脉冲滤波算法 (FIR)、无限脉冲滤波 (IIR)、自相关 (Autocor) 和差的绝对值和 (SAD) 等典型算法, 针对利用 VM 的非对齐访问和只利用 VM 的对齐访问功能两种情况进行编程测试, 分析 YHFT-Matrix 完成向量运算所需的访存和混洗指令数, 结果如图 7 所示。

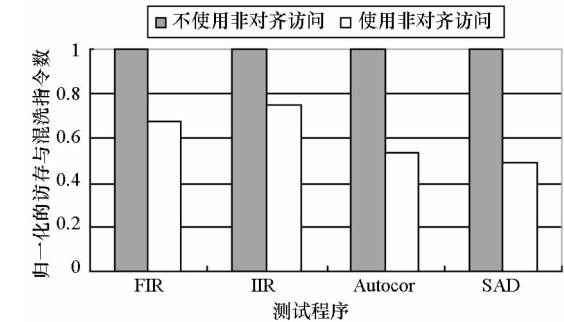


图 7 不使用 VM 对齐和使用 VM 对齐功能的性能对比
Fig. 7 The performance compare between the unaligned VM and aligned VM

其中 FIR、IIR 均为 1024 点、16 抽头的 32 位样本数据, Autocor 为 512 点、8 抽头的 16 位样本数据, SAD 为实时图为 256×256 像素、模板为 64×64 像素的 8 位数据。所有算法所需源操作数均为定点数, 采用手工编写的经过优化的汇编代码实现。

从图 7 中可见, 对于滤波、自相关一类算法, 由于运算时需要大量向量访存地址跨步为 1 的连续向量数据, 使用 VM 的非对齐访问功能后, 上述 4 个测试程序的向量访存和混洗指令数之和将降低 31%、25%、47% 和 49%, 具有较高的访存效率。

因此, 基于多路 SIMD 结构所设计的 VM 不仅能支持多端口并行访存, 有效降低访存冲突, 还支持非对齐方式的向量数据访问, 对于需要大量的向量访存地址跨步为 1 的向量访存的一类算法, 实现了多路运算单元对向量存储器的共享, 大幅减少或消除了相关算法的混洗操作, 压缩了代码密度, 从而提高了访存效率, 加速了相关算法。

参考文献 (References)

[1] Lin Y, Lee H, Woh M, et al. SODA: A High-Performance DSP architecture for software-defined radio[J]. IEEE Micro, 2007(1): 114-122.

[2] Cho J, Chang H, Sung W Y. An FPGA based SIMD processor with a vector memory unit[C]// ISCAS 2006: 525-528.

[3] Talla D, John L K, Burger D. Bottlenecks in multimedia processing with SIMD style extension and architectural enhancements[J]. IEEE Transactions. Computers, 2003, 42(8): 1015-1031.

[4] Khailany B, Dally W J, Chang A, et al. VLSI design and verification of the imagine processor[C]// Proceedings of the IEEE International Conference on Computer Design, 2002: 289-296.

[5] Chang H, Cho J, Sung W Y. Performance evaluation of an SIMD architecture with a multi-bank vector memory unit[C]// IEEE workshop, SIPS'2006: 71-76.

[6] Wob M, Seo S W, Mahlke S, et al. AnySP: anytime anywhere anyway signal processing[C]// ISCA'2009: 128-139.

一种面向语义正确性的关系数据库访问方法^{*}

唐富年,姚莉,李金洋

(国防科技大学 信息系统与管理学院,湖南 长沙 410073)

摘要:从异构关系数据源中获取语义正确的信息是当前利用数据资产辅助决策所面临的一大难题。为解决这一问题,提出了一种研究面向语义正确性的关系数据库访问方法(SCORDA)。SCORDA方法采用DL-Lite_{OWL}^N本体作为支持数据访问任务的概念视图,在本体的表达能力和推理复杂度之间做了一定折中;SCORDA方法通过LAV+GAV的映射方式建立本体与关系数据源之间的语义关联,兼顾了模式异构和阻抗失配问题;此外,SCORDA方法采用了一种动态ABox结构,即时从异构关系数据源中恢复虚拟对象,并且利用动态ABox进行实例检测推理任务,通过本体的模型语义来保证数据访问的正确性。通过原型系统验证了该方法的可行性。

关键词:本体;关系数据库;语义Web;数据语义

中图分类号:TP182 **文献标志码:**A **文章编号:**1001-2486(2012)03-0103-06

A semantic correctness oriented relational database access method

TANG Funian YAO Li LI Jinyang

(College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

Abstract: Nowadays, organization faces the problem of acquiring correct information from data asset to support decision-making. To solve this problem, a semantic correctness oriented relational data access method, namely SCORDA, is proposed. Firstly, ontology was used to describe the concept view, which is used in data access task, and description logic makes a suitable tradeoff between expressiveness and complexity. Secondly, a LAV+GAV mapping formalism was adopted to solve the schema heterogeneity and impedance mismatch problem. Furthermore, a special structure named dynamic ABox is provided to deal with the instance checking problem, and to ensure the semantic correctness of answers based on the model theory semantic of ontology. Finally, the feasibility of SCORDA method, through implementing a prototype system, was validated.

Key words: ontology; relational database; semantic Web; data semantic

数据是信息时代最重要的战略资源之一,关系数据库存储的数据是目前世界上最主要的数据来源^[1]。然而,尽管人类社会的数据资产逐年递增,但是如何从中访问到语义正确的信息,将信息优势转化为决策优势,仍然是当前很多组织面临的难题。

为解决上述问题,近年来基于本体的数据访问技术开始得到人们的关注^[2]。为了保证数据访问结果的语义正确性,通常需要从数据模型固有的内涵语义出发来认知和利用数据^[3]。其主要技术难点在于:

(1)由于关系模型和本体模型之间存在阻抗匹配问题^[4],必须采用较为复杂的语义映射,而这种映射的构造通常困难而繁琐。

(2)由于现有推理机只能支持中等规模

ABox的推理,传统的推理算法和推理机都难以直接应用于大规模的数据访问^[5]。

(3)在动态环境下,当本体或者关系数据库发生演化时,需要确保关系数据库和本体之间的映射保持正确,进而保证数据访问的语义正确性^[2]。

为此,本文提出了一种面向语义正确性的关系数据库访问(SCORDA, Semantic Correctness Oriented Relational Data Access,)方法。该方法采用DL-Lite_{OWL}^N本体作为支持语义访问的概念视图,通过LAV+GAV形式的映射方式建立关系数据库和本体之间的语义关联,借助动态ABox机制完成查询的处理。最后,通过原型系统和相关实验证明该方法是可行的。

^{*} 收稿日期:2011-12-20

基金项目:国家自然科学基金资助项目(70971134)

作者简介:唐富年(1981-),男,内蒙古阿拉善盟左旗人,博士研究生,E-mail:tfn1981@gmail.com;

姚莉(通信作者),女,教授,博士,博士生导师,E-mail:Liyao6522@sina.com

1 基本框架与形式化描述

关系数据库的语义在概念建模阶段确定,而在数据库运行阶段,这些数据语义却经常难以被一般用户理解和重用。SCORDA 方法的基本思路是通过一个本体描述的概念视图将底层的异构数据源集成到本体知识表示模型的架构之下,借助知识库的知识表示和推理为上层应用提供语义正确的信息。采用 SCORDA 方法构建的信息系统可以如下定义。

定义 1 SCORDA 系统可定义为 $K = \langle E, O, DB, M \rangle$, 其中: E 为系统的论域, O 为一个能够充分描述关系数据库模式语义的本体, DB 是 E 上的关系数据库集合。 M 是关系数据库与本体之间的语义映射集合, $M = M_L \cup M_C$ 。 其中, M_L 为关系数据库到本体的 LAV 映射, 而 M_C 为基于 M_L 构造的 GAV 映射。

本体的选择是 SCORDA 方法关键的环节, 直接决定了整个系统的性能。为了兼顾数据规模、本体表达能力和查询处理方面的需求, 本文在 DL-Lite 系列描述逻辑子集^[6]的基础上进行扩展。DL-Lite_A 描述逻辑子集是 DL-Lite 系列中面向数据访问设计的一个子集, 但是该子集牺牲了数量约束等对于关系数据库建模较为重要的约束。根据文献[7], 将某些 OWL 中的标准构造子加入 DL-Lite 子集后不会不影响其计算特性, 例如属性的对称性和反自反性约束等; 在命名唯一假设(unique name assumption)和无角色包含的情况下, 引入数量约束(number restriction)也不会增加原描述逻辑的计算复杂度。

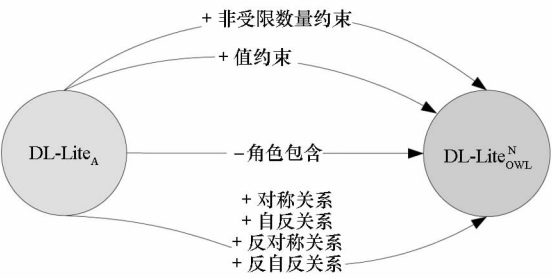


图 1 从 DL-Lite_A 到 DL-Lite^N_{OWL}

Fig. 1 From DL-Lite_A to DL-Lite^N_{OWL}

DL-Lite^N_{OWL} 描述逻辑子集兼顾了 DL-Lite_A 和 OWL 本体语言中的一些重要特性, 同时数据复杂度仍然保持了较低水平。

定义 2 DL-Lite^N_{OWL} 本体可以定义为一个三元组 $O = \langle \Gamma, T, A \rangle$, 其中:

(1) 符号集 $\Gamma = C \cup Op \cup Dp \cup V$, 其中, 概念 C

$= T_c \mid B \mid \neg B \mid \delta(Dp) \mid \exists Op. C \mid \forall Op. C$ (T_c 为顶概念, B 为原子概念, $\delta(Dp)$ 为 Dp 的定义域); 对象属性集合 $Op = C \times C$; 数据类型属性集合 $Dp = C \times V$ 。

(2) TBox T 是一个有限的公理集合, 公理类型包括以下几种:

① 概念包含公理: $C_1 \subseteq C_2$ 。

② 属性约束公理: 包括函数性公理 $func_t(Op)$ 和 $func_t(Dp)$ 、对称性公理 $sym(Op)$ 、反对称性公理 $asym(Op)$ 、自反公理 $ref(Op)$ 和反自反公理 $irr(Op)$ 。

- ABox A 为一个有限的断言集合。

定义 3 对于给定的 DL-Lite^N_{OWL} 本体 O , O 的解释为 $I = (\Delta^I, \cdot^I)$, 其中:

(1) Δ^I 为解释域, Δ^I 由对象域 Δ'_o 和取值域 Δ'_v 两个互不相交的部分组成, 即 $\Delta^I = \Delta'_o \cup \Delta'_v$ 。

(2) \cdot^I 为解释函数, 该函数为本体 O 中的公理和断言赋予形式化的含义。

为了便于进行查询的处理, 将知识库的 TBox T 划分为仅含有肯定包含公理的 T_d 和含有其他约束公理的 T_{cst} 两个部分。同时, 为了适应这种划分又采用一种动态 ABox 结构, SCORDA 系统的基本结构如图 2 所示。当进行查询应答时, 首先基于 T_d 抽取与用户查询相关的对象信息动态形成一个 ABox (即动态 ABox), 随后通过本体推理进一步求精, 返回满足全部公理约束的应答。

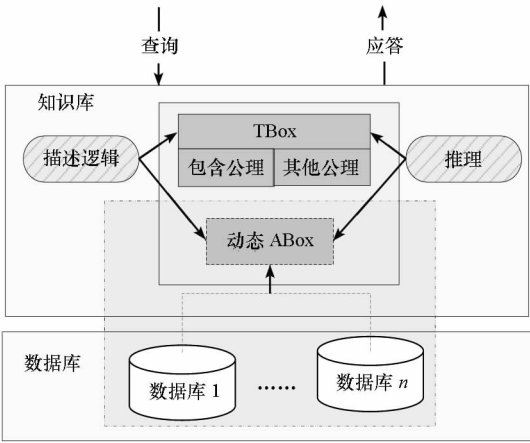


图 2 SCORDA 的知识表示模型

Fig. 2 Knowledge representation model of SCORDA

动态 ABox 仅在查询应答过程中出现, 当查询任务结束, 则 ABox 也随之释放。采用动态 ABox 的 SCORDA 系统呈 3 种状态:

(1) 由 T_d 和 A_ϕ 构成的初始状态, 记作 K_0 。

(2) 由 T_d 和 A_d 构成的第一阶段查询处理状态, 记作 K_1 。其中 A_d 中的断言为针对 T_d 中的公理获得的结果集。

(3) 由 $T_d \cup T_{cst}$ 和 A_d 构成的第二阶段查询处

理状态,记作 K_2 。

在 K_0 状态下,DL-Lite_{OWL}^N 知识库仅由 T_D 组成,由于 T_D 中仅包含肯定包含公理,此时知识库 $\langle T_D, A_\phi \rangle$ 中不存在不一致;在 K_1 状态下, A_D 已经非空,此时知识库中 $\langle T_D, A_D \rangle$ 仍然不存在不一致。在 K_2 状态下,知识库 $\langle T_D \cup T_{Cst}, A_D \rangle$ 可能存在不一致,此时需要对 A_D 中的个体进行筛选。

定义4 对于 SCORDA 系统的任意查询 q , 满足 q 的确定应答 $CAns(q, O)$ 是动态 $ABox$ 中的一个常量元组 t , 且 t 中包含 VO 中对象的全部或部分信息,对于本体 O 的每一个模型 $Mod(O)$, 都有 $t \in q^I$ 。

如果 O 的一个解释 I 能够满足 O 中的所有公理和断言,则 I 为 O 的一个模型。为了保证数据访问的语义正确性,为本体找到模型是进行数据访问的关键步骤,只有满足本体模型语义的数据才能够提交给用户,这也是确保数据语义正确性的必要环节。

2 映射的构建

目前,关系数据库和本体之间的映射并没有统一的表达形式,根据具体应用的不同,映射所含有语义信息的量级也有所不同。为了克服关系数据库模式和本体间固有的阻抗失配问题,通常需要使用 GAV、LAV 和 GLAV 等形式的复杂映射表达式^[3,8-9]。

SCORDA 方法首先需要构建关系数据库到本体的 LAV 映射,然后基于该 LAV 映射自动构造 GAV 映射。其中,LAV 映射的构造过程本质上是对关系数据库语义的挖掘和描述,使用逻辑公式表达出每个关系模式的语义;GAV 映射的构造则将本体论域上的常量对象和关系数据库中的元组对应起来。而且,从关系数据库到本体的 LAV 映射可以借助映射发现算法以半自动化的方式生成,而 GAV 映射的构造过程可以基于 LAV 映射和相应的算法自动完成。这种结构最大的优点在于整合了 LAV 和 GAV 的优势,每当有新的关系数据源加入或者本体发生演化时,系统能够很快自适应。

2.1 LAV 映射的发现

LAV 映射本质上就是将关系数据库模式上关系数据库模式的语义准确“投射”到本体之上,使用本体中的元素将隐含于关系数据库模式中的语义显式表示出来。针对 DL-Lite_{OWL}^N 本体的基本特点,本文采用一种基于概念连接图的映射发现

方法。

定义5 关系数据库模式 S 到本体 O 的 LAV 映射 M_L 定义为一个形如 $T(X) \rightarrow \Psi(X, Y)$ 的一阶逻辑表达式。其中, $T(X)$ 表示 S 中的一个关系模式; $\Psi(X, Y)$ 表示本体上的一个合取公式,所有的谓词均来自 O ; 变量集合 X 代表关系模式中的列,而变量集合 Y 代表 DL-Lite_{OWL}^N 本体论域上的对象。

定义6 如果 DL-Lite_{OWL}^N 本体中的某一元素 $e \in C \cup Op \cup V \cup Dp$ 在公理 $\alpha \in T_D$ 和 $\beta \in T_D$ 中都出现,则称公理 α 和公理 β 关于 e 结构连接。

定义7 DL-Lite_{OWL}^N 本体 T_D 中的公理构成了一个概念连接图 $G_C = \langle CNodes, Edges \rangle$, 其中图的节点均为本体中的概念,图的边可以为包含公理或者结构连接关系。

假设有如下的 DL-Lite_{OWL}^N 本体片段,则相应的概念连接图片段如图3所示,图中

Conference \subseteq Event \exists OrgnizerOf \subseteq Event
 Person $\subseteq \exists$ OrganizerOf participantOf \subseteq Event
 Person $\subseteq \exists$ participantOf $\rho(\text{hasStartTime}) \subseteq$ Event
 Chair \subseteq Person $\rho(\text{endTime}) \subseteq$ Event
 $\rho(\text{name}) \subseteq$ Person $\rho(\text{eventTitle}) \subseteq$ Event

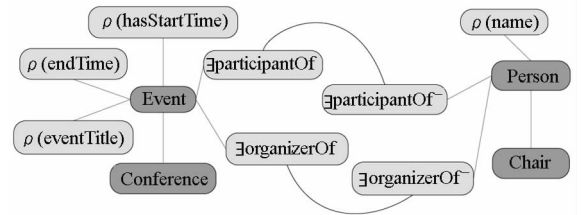


图3 概念连接图片段示例

Fig. 3 Example of concept connection graph

基于概念连接图发现 LAV 映射的基本步骤为:

(1) 建立关系数据库模式中的符号到本体中的符号的对应。假设有前述示例公理集和关系模式 **chair** (name, eventTitle, startingTime, endTime), 则可以添加对应:

DB. **chair**. name \leftrightarrow Onto. Chair. name

DB. **chair**. eventTitle \leftrightarrow Onto. Conference. eventTitle

DB. **chair**. startingTime \leftrightarrow Onto. Conference. hasStartTime

尽管当前模式匹配算法在精度上有了较大提升,但是仍然无法确保匹配结果的完全正确,通常需要人工参与对应构建过程。后文中假设已经获得了符合要求的对应。

(2) 标识关系模式对应的概念。概念的标识利用主外键约束和上一步给定的对应集合实现,例如 **chair** 表反映了多对多关系,根据用户输入

的对应,可以标识出隐含实体集 *Chair* 和 *Conference*。相关内容可参阅文献[10–11]。

(3) 基于概念连接图进行最小公理集搜索。当关系模式标识出多个概念时,为了形成闭合的语义关系,需要从本体中抽取一个连接这些概念的最小公理集。例如在 $DL-Lite_{OWL}^N$ 中,概念 *Chair* 和 *Conference* 之间的最小公理集为:

$Chair \subseteq Person \quad \exists OrganizerOf \subseteq Person$
 $\exists OrganizerOf^- \subseteq Event \quad Conference \subseteq Event$

(4) 将一阶逻辑映射公式编码。编码一阶逻辑形式的 LAV 映射。例如:

$chair(name, eventTitle, startingTime, endTime) \rightarrow Chair(x_1) \wedge$
 $Person(x_2) \wedge isa(x_1, x_2) \wedge name(x_2, name) \wedge$
 $Conference(y_1) \wedge Event(y_2) \wedge isa(y_1, y_2) \wedge$
 $eventTitle(y_2, eventTitle) \wedge hasStartTime(y_2, startingTime) \wedge$
 $hasEndTime(y_2, endTime) \wedge OrganizerOf(x_2, y_2)$

由于存在“SubClassOf”关系的概念之间存在属性的继承关系,因此上述表达式也可以进行语义关系的合并,进一步简化为如下形式:

$chair(name, eventTitle, startingTime, endTime) \rightarrow Chair(x) \wedge$
 $name(x, name) \wedge Conference(y) \wedge$
 $hasStartTime(y, startingTime) \wedge$
 $hasEndTime(y, endTime) \wedge OrganizerOf(x, y)$

2.2 GAV 映射的构造

SCORDA 方法的基本思想就是借助关系数据库和本体间的映射从关系数据库恢复常量对象集,进而利用对象间的语义关系提高查询的语义正确性。从 LAV 映射生成 GAV 映射的常用方法是引入 Skolem 函数,根据键约束生成对象标识符。

(1) 涉及单个概念的 LAV 映射。假设已有如下 LAV 映射(关系表中未参与映射的列记作“_”):

$person(name, _, _, email) \rightarrow Person(x) \wedge$
 $name(x, name) \wedge email(x, email)$

假设 Skolem 函数 ff , 可以得到如下 GAV 映射:

$Person(ff(name), email(ff(name), email)) \rightarrow$
 $person(name, _, _, email)$

(2) 涉及两个概念及概念间关系的 LAV 映射。以前述 **chair** 表形成的 LAV 映射为例,假设 Skolem 函数 f 和 g , 则可以恢复出如下 GAV 映射:
 $Chair(f(name)) \rightarrow chair(name, _, _, _)$
 $Event(g(eventTitle), hasStartTime(g(eventTitle),$
 $startingTime), hasEndTime(g(eventTitle), endTime))$
 $\rightarrow chair(_, eventTitle, startingTime, endTime)$
 $OrganizerOf(f(name), g(eventTitle)) \rightarrow$

$chair(name, eventTitle, _, _)$

由于 *chair* 表中的外键 *name* 参考了 *person* 表中的主键 *name*, 为了恢复出 *chair* 对象集的完整信息,还需要对 *chair* 表和 *person* 表进行右外连接:
 $Chair(f(name), email(ff(name), email)) \rightarrow$

$\Pi_{person.name, person.email}(chair(name, _, _, _))$
 $person(name, _, email))$

(3) 涉及三个以上概念的 LAV 映射的处理与(2)类似,只是需要进行多个对象集的恢复。

由于在 LAV 映射构造过程中已经对数据库语义进行了正确解释,上述映射转换可以自动完成,节省了较多工作量。在具体实现时, GAV 映射右端通常是一个 SQL 语句,前述映射可以写作:

$Chair(f(name), email(ff(name), email)) \rightarrow$
 $select name, email from person right outer$
 $join chair on person.name = chair.name$

3 查询处理及实例检验

基于 SCORDA 方法构造的信息系统能够应答合取查询(Conjunctive Query, CQ)、合取查询的并(Union of Conjunctive Query, UCQ)和布尔查询(Boolean Query, BQ)。其中, CQ 是最基本的查询表达形式,它对应关系数据库的 SPJ (Select Project Join) 查询,能够满足常规用户的数据访问需求,其定义如下:

定义 8 $DL-Lite_{OWL}^N$ 本体上的合取查询是形如:

$$Q(x) : - \bigwedge_{0 < i}^n Atom_i(x, y)$$

的表达式,其中,合取项 $Atom_i$ 可以为 $C(x)$ 、 $Op(x, y)$ 、 $Dp(x, v)$ 等形式的原子, $x, y \in \Delta_o^I$ 且 $v \in \Delta_v^I$ 。此外,出现在查询头部的变量为 distinguished 变量,在多个原子中都出现的变量为共享变量。

例如查询 $Q(x) : - paper(x) \wedge hasAuthor(x, z)$ 的语义就是查找一个属于概念 *paper* 的对象,并且它的 *hasAuthor* 属性为任意值。

UCQ 查询是形如 $Q(x) : - CQ_1 \vee CQ_2 \vee \dots$ 的查询, UCQ 查询的应答集等价于各析取项(CQ)应答集的并。布尔查询是形如 $Q : - paper(x) \wedge hasAuthor(x, z)$ 的查询,这类查询并不返回对象,仅返回布尔值,其用途为询问知识库中是否存在查询表达式所描述的事实。

考虑到 $DL-Lite_{OWL}^N$ 本体遵循命名唯一性假设,要采用常规 DL 推理机进行实例检验就必须考虑对象 IRI 的动态分配问题。通过 GAV 映射

表达的每一个对象集的 URI 为“本体 URI + 数据库名 + 对象集标识名”,这样即可确保标识符的唯一性。参考文献[12]给出如下定义:

定义 9 虚拟对象模式可以定义为一个三元组 $VScheme = \langle Cls, Attri, URI \rangle$, 其中: Cls 为虚拟对象的类名集合, $Attri$ 为虚拟对象的特性集合, URI 为标识虚拟对象的 URI。

定义 10 虚拟对象模式的数据库实例定义为 $J = \langle VO, M_C, \mu \rangle$, 其中, VO 为一个有限的虚拟对象集合, 映射 M 将 DB 中的某一元组指派给 VO 中某一对象, 函数 μ 为 VO 的每一个虚拟对象分配一个不同的 URI。

虚拟对象并非 Δ_o^I 上实际存在的对象, 在数据访问过程由关系数据库中的元组动态填充生成。查询处理过程分为 3 个阶段:

(1) 用户查询的扩展重写阶段。前文已经提到, SCORDA 知识库的 TBox 部分由 T_D 和 T_{Cst} 两部分组成。在查询处理过程中 T_D 中的公理首先参与用户查询 q 的扩展重写环节, 得到重写后的 UCQ 查询 q' 。该过程可参见文献[6]中的 PerfectRef 算法。

(2) 动态 ABox 构造阶段。如图 4, 查询 q' 对应的是一个对象集合, 而该对象集合是通过数据库中的数据来构造的。此时, 需要基于查询 q' 和 GAV 映射从关系数据库中获取相关的数据, 为每个对象集分配 IRI, 构成动态 ABox。

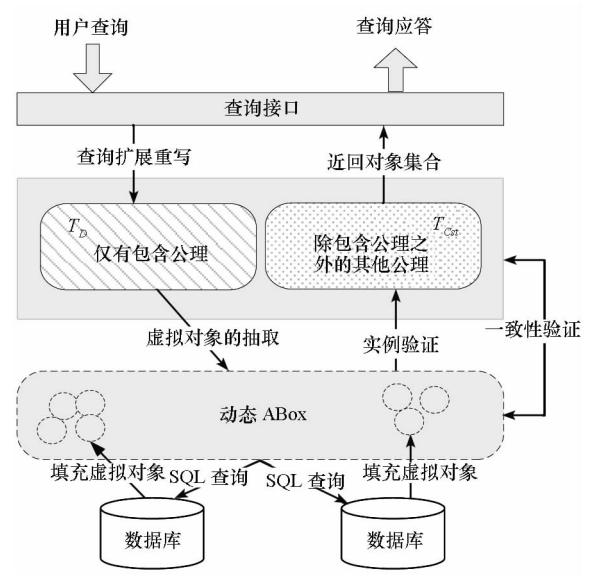


图 4 查询处理流程
Fig. 4 The flow of query processing

(3) 实例检测阶段。由于查询的模式重写过程仅仅使用了 TBox 中的部分公理, 因此无法确保获取到的应答集能够满足所有的公理约束, 即无法保证当前知识库一致。实例检验过程的任务是

通过本体 ABox 推理将不符合 T_{Cst} 中公理约束的对象筛选掉。与 OBDA 方法等相比, SCORDA 方法可以采用常规 DL 推理机进行实例检验, 具有更好的可伸缩性。

4 原型系统实现

SCORDA 系统可分为映射管理模块和数据访问模块两部分分别予以实现。其中, SMMS (Semantic Mapping Management System) 映射管理模块如图 5 所示, 在实现过程中重用了 MAPONTO^[11] 系统中关系数据库模式操作的部分代码, 详细的映射发现算法参见文献[10]。



图 5 SMMS 语义映射构造系统
Fig. 5 Semantic mapping management system

在数据访问模块的实现过程中使用 MAPONTO 实验数据集中的 VLDB 数据库和 D2RQ^[13] 示例中的 ISWC 数据库为数据源, 采用 MAPONTO 实验数据集中的 conference. owl 本体作为概念视图 (精简为 DL-Lite_{OWL}^N 本体)。本体操作和推理通过 OWL API 和 Pellet 推理机完成。例如, 假设用户查询为 $Q(x, y) : \neg hasLocation(x, y)$, 则系统将返回如图 6 所示的结果, 图中的右半部分界面给出了本体的继承结构, 便于用户构造查询。



图 6 数据访问模块
Fig. 6 Data access module

SCORDA 方法利用关系数据库模式的内涵语义进行数据访问,并且通过本体的模型论语义保证数据访问的正确性,因而数据访问的语义正确性是有保证的。限于文章篇幅,本文中仅对 SCORDA 方法的基本框架、查询处理流程和可行性等进行了阐述和验证,关于数据规模和查询响应时间方面的性能比较和论证将在后续文章中给出。

5 结束语

为了促进信息优势向决策优势转换,本文提出了一种面向语义正确性的关系数据库访问方法 SCORDA,给出了完整的解决方案并且验证了其可行性。与已有的 OBDA 等方法相比,该方法对 DBMS 和数据联合工具的依赖度大大下降,并且能够借助常规 DL 推理机完成实例检测推理任务。下一步的工作将就不同数据规模下 SCORDA 方法的稳定性能进行研究,对原型系统做进一步完善。

参考文献 (References)

[1] Chang K C, He B, Li C, et al. Structured databases on the Web: observations and implications [J]. SIGMOD Record. 2004, 33(3): 61 – 70.

[2] 唐富年, 姚莉, 李金洋. 基于本体的关系数据库访问研究进展[J]. 小型微型计算机系统, 2011, 32(3): 390 – 396.

TANG Funian, YAO Li, LI Jingyang. Ontology-based relational database access: a survey and future direction[J]. Journal of Chinese Computer System, 2011, 32(3): 390 – 396. (in

Chinese)

[3] Borgida A, Mylopoulos J. Data semantics revisited [C]// Proceedings of the Workshop on Semantic Web and Databases (SWDB) in Conjunction with the International Conference on Very Large Data Bases (VLDB), 2004: 9 – 26.

[4] Poggi A, Lembo D, Calvanese D, et al. Linking data to ontologies [J]. Journal on Data Semantics X. 2008: 133 – 173.

[5] Rodriguez-muro M, Calvanese D. Dependencies: making ontology based data access work in practice [C]// Proceedings of the 5th Alberto Mendelzon International Workshop on Foundations of Data Management. Santiago, Chile: 2011.

[6] Calvanese D, de Giacomo G, Lembo D, et al. Tractable reasoning and efficient query answering in description logics: the DL-Lite family [J]. Journal of Automated Reasoning, 2007.

[7] Artale A, Calvanese D, Kontchakov R, et al. The DL-Lite family and relations [J]. Journal of Artificial Intelligence Research, 2009, 36: 1 – 69.

[8] Calvanese D, de Giacomo G, Lembo D, et al. Using OWL in data integration [M]. Semantic Web Information Management—A Model Based Perspective, Springer, 2009.

[9] Poggi A, Rodriguez-muro M, Ruzzi M. Ontology-based database access with DIG-mastro and the OBDA plugin for protégé [C]// Proceeding of the Workshop OWLED. 2008.

[10] Tang F N. Discovering semantic mappings for ontology-based data access [C]// ICDDM 2011, SanYa, China: 2011.

[11] An Y. Discovering and using semantics for database schemas [D]. Toronto: University of Toronto, 2007.

[12] Calvanese D, Lenzerini M, Nardi D. Unifying class-based representation formalisms [J]. Journal of Artificial Intelligence Research, 1999, 11(2): 199 – 240.

[13] Bizer C, Seaborne A. D2RQ-treating non-RDF databases as virtual RDF graphs [C]// The 3rd International Semantic Web Conference, 2004.

警报关联图：一种网络脆弱性量化评估的新方法^{*}

张 怡, 赵 凯, 来 犇

(国防科技大学 计算机学院, 湖南 长沙 410073)

摘 要: 作为一种基于模型的脆弱性分析技术,攻击图能够识别网络中存在的脆弱性和它们之间的相互关系,分析出可能的攻击路径和潜在威胁。论文在攻击图的基础上提出了警报关联图的概念,利用攻击图中蕴含的脆弱性先验知识,将实时 IDS 警报信息映射到攻击路径,动态反映攻击进程和攻击者意图。在此基础上提出了一种基于警报关联图的网络脆弱性量化评估方法,通过计算警报关联边的权值对网络脆弱性进行动态分析,这种方法结合了静态的网络脆弱性先验知识和动态变化的攻击者意图,能有效反映网络脆弱性在动态攻击情况下的变化。

关键词: 攻击图;警报关联图;脆弱性评估

中图分类号: TP393.08 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0109-04

Alert correlation graph: a novel method for
quantitative vulnerability assessment

ZHANG Yi, ZHAO Kai, LAI Ben

(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: As a model-based vulnerability analysis technology, attack graphs can identify network vulnerabilities and their interactions; they can also reveal all possible attack paths and potential threats. Based on the attack graphs, alert correlation graphs are proposed in the paper. An alert correlation graph maps real-time IDS alerts into attack paths using prior knowledge encoded in attack graph, and reveals attack progresses and attackers' intention dynamically. A novel quantitative network vulnerability assessment method is presented based on the alert correlation graph, which analyzes network vulnerabilities by dynamically computing the weight of alert correlation edges. The research also demonstrates, by examples, that the proposed method combines static prior knowledge about network vulnerabilities with dynamic attackers' intentions, and reveals the change of network vulnerability under real-time attacks.

Key words: attack graph; alert correlation graph; vulnerability assessment

网络脆弱性是影响网络安全的关键因素。网络中的脆弱性表现在设计、实现和运行管理的各个环节^[1]。完全消除脆弱性是不现实的,也是不可能的,重点是对网络脆弱性进行分析和度量,找出影响网络的关键脆弱性,有的放矢地进行处理。攻击图是基于模型的网络脆弱性分析技术,它从攻击者的角度出发,在综合分析网络配置和脆弱性相关信息的基础上,枚举出所有可能的攻击路径,反应了网络中各个脆弱性之间,以及脆弱性与安全配置之间的相互依赖和相互作用关系,能够深层次地揭示网络中存在的潜在威胁。攻击图技术已被证明在网络脆弱性分析、网络加固、攻击响应等研究领域具有良好的应用前景。

目前,攻击图建模和自动生成技术研究取得较大进展^[2-4],越来越多的研究集中在基于攻击图的网络脆弱性分析技术上,即根据攻击图中节

点的实际含义,采用各种模型或算法对网络脆弱性进行分析。如文献[5]为攻击图中每个原子攻击指派成功发生的概率值,利用马尔科夫模型计算攻击目标被攻击者成功入侵的最大概率。文献[6]通过配置网络中关键信息资产的价值,对网络的脆弱性进行度量。文献[7]利用专家经验确定攻击图中每个节点独立发生的概率,并计算从攻击初始节点到达每个属性节点的累计概率。上述各种方法中,对攻击图中节点或边的参数分配依赖于管理人员的经验,主观性强,可操作性较差。考虑到 IDS (Intrusion Detection System) 警报反映了网络中客观存在的攻击行为,文献[8]中首次将攻击图与 IDS 警报关联相结合,提出了基于队列图(queue graphs)结构的警报关联计算方法,不但提高了警报关联计算的效率,而且还基于攻击图信息,将警报关联(correlation)、假设

^{*} 收稿日期:2011-09-20
基金项目:国家 863 计划资助项目
作者简介:张怡(1973—),女,山西五台人,副研究员,博士,硕士生导师,E-mail:zhangyi@nudt.edu.cn

(hypothesis) 和预测(prediction)进行了统一。

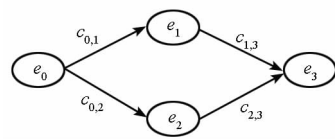
受队列图的研究启发,本文提出警报关联图思想。警报关联图不但包含了 IDS 警报间的关联关系,而且还对这些关联发生的次数进行量化。由于攻击者能主动分析各种因素,智能地进行攻击规划,因此警报关联图可真实反映攻击者意图,可用来对网络脆弱性进行动态的评估。本文提出的方法可应用于分析攻击规律、提取攻击模式、蜜网分析、安全策略验证和部署调整以及使用历史警报数据对攻击图进行预先模拟分析等。与其他分析方法相结合,警报关联图可以为进一步的网络脆弱性分析提供基础数据,指导确定网络防护措施。

1 警报关联图

警报关联图的基本思想是在攻击图提供的先验知识基础上,根据 IDS 警报信息动态生成警报关联,并基于警报关联的次数计算关联边的权值。

图 1(a)所示为简单的攻击图。其中 e_0 、 e_1 、 e_2 和 e_3 为表示原子攻击的利用节点, v_0 、 v_1 和 v_2 为表示利用条件和后果的属性节点。 e_0 攻击后进入 v_0 状态,在 v_0 状态分别可通过实施 e_1 和 e_2 攻击进入 v_1 和 v_2 状态,而只有 v_1 和 v_2 状态同时达到时,攻击者才可以实施 e_3 攻击。文献[9]中给出了攻击图的形式化定义。图 1(b)是图 1(a)对应的警报关联图,其节点集合是攻击图中利用节点(原子攻击节点)集合,有向边反映了利用节点对应的 IDS 警报之间的关联关系,边的权值(如 $c_{0,1}$)代表警报间的关联次数。显然,警报关联图的有向边也反映了攻击路径,边的权值越大说明攻击者沿着该路径进行攻击的次数越多。

基于攻击图可对 IDS 的警报进行关联。由于 IDS 的警报信息是相互独立的,并且可能存在缺失,因此需要对缺失的警报进行推断。文献[8]将攻击图的属性节点分为 TRUE、FALSE 和 HYP 三个状态。TRUE 表示该节点的属性得到满足,可以根据该属性实施下一步的攻击。FALSE 状态表示该节点的属性不满足,无法实施后续攻击。而 HYP 节点代表根据推断可以假设该节点属性已经满足,但还没有相应的警报信息证实。本文采用同样的属性状态分析警报关联图。



(b) 警报关联图

图 1 警报关联的基本原理

Fig. 1 Principle of alert correlation

在图 1(a)中,假设在 t_1 时刻系统得到对应原子攻击 e_0 的警报 a_0 ,由于攻击 e_0 直接导致属性 v_0 可达,因此警报 a_0 触发将 v_0 状态改为 TRUE。假设 t_2 时刻对应 e_3 的警报 a_3 出现,说明攻击者已经获得 v_1 和 v_2 属性,由于此时系统尚未收到警报 a_1 和 a_2 ,因此只能将 v_1 和 v_2 的状态设置为 HYP,即可推断 e_1 和 e_2 发生,造成属性 v_1 和 v_2 的满足。而当 t_3 时刻警报 a_2 出现时, v_2 的状态由 HYP 变为 TRUE。在上述 IDS 警报序列下,属性节点的状态变化以及警报关联图边权值的变化如图 2 所示。警报关联图边权值的计算方法将在下一节中详细介绍。

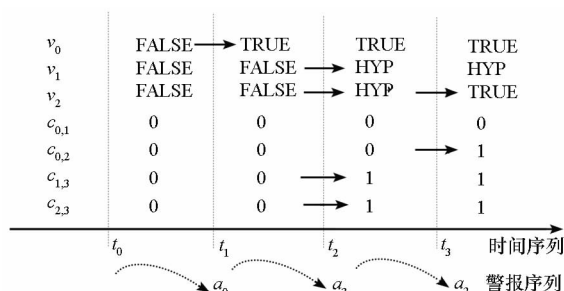


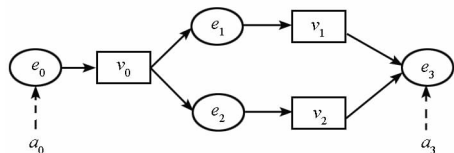
图 2 警报关联中的状态与权值变化

Fig. 2 Change of state and weights in alert correlation

因此,警报关联图可定义为三元组 $\langle V, E, C \rangle$,其中顶点集合 $V = \{v_i | i = 0, 1, \dots, n-1\}$,对应攻击图中的利用节点集合,边集合 $E \subseteq V \times V$ 表示警报之间的关联关系,关联度集合 $C = \{c_{i,j} | v_i, v_j \in V\}$,定义了根据 IDS 警报动态计算的关联边的警报关联次数。

2 警报关联度计算

警报关联度计算是指当一个 IDS 警报发生时,根据攻击图(队列图)当前状态和警报关联结果,修改警报关联图中相应边的关联度的过程。根据攻击图的定义^[9],属性节点的父节点(利用节点)之间为或关系,表示父节点中任何一个原子攻击实施成功,都会使该属性节点被满足;利用节点的父节点(属性节点)之间为与关系,表示利用节点的全部前提属性必须都被满足,才能实施相应的原子攻击。因此,计算警报关联度需要考虑图 3 中(a)、(b)、(c)3 种情况,其中实心矩形



(a) 攻击图

表示属性节点,空心圆形表示利用节点, e_0 为当前警报对应的利用节点,其余为 e_0 的前续利用节点,也是在警报关联图中具有出边连接 e_0 的节点。

图3中,对于情况(a), e_0 攻击发生,必然有 e_1 攻击先发生,因此 e_0 对应的警报发生时,将 $c_{1,0}$ 加1;对于情况(b), e_0 攻击发生,必然有 e_1 和 e_2 攻击先发生,因此 e_0 对应的警报发生时,将 $c_{1,0}$ 和 $c_{2,0}$ 分别加1;而对于情况(c), e_1 和 e_2 攻击有一个发生,就可能使 e_0 的前续属性节点状态变为 TRUE,因此,当 e_0 警报发生时,无法直接推断其前续攻击是 e_1 还是 e_2 ,此时采用的计算方法是若 e_1 对应的警报存在,而 e_2 对应的警报不存在, $c_{1,0}$ 加1, $c_{2,0}$ 不变,反之, $c_{2,0}$ 加1, $c_{1,0}$ 不变;若 e_1 和 e_2 对应的警报都存在或都不存在,则 $c_{1,0}$ 和 $c_{2,0}$ 各加0.5。当然,在警报关联计算时需要考虑(a)、(b)、(c)3种情况混合存在的情况,如图3(d)所示。

为实现警报关联计算,定义 $N \times N$ 的关联度矩阵 C 和警报状态向量 S ,其中 N 为警报关联图中的节点个数, $c_{i,j}$ 表示节点 e_i 和 e_j 之间的关联度。 s_i 为1表示 e_i 对应的警报已经出现,为0表示未出现。对于每个利用节点,都对应一个用于警报关联计算的数据结构。例如,图3(d)中节点 e_0 对应的数据结构如图3(e)所示。

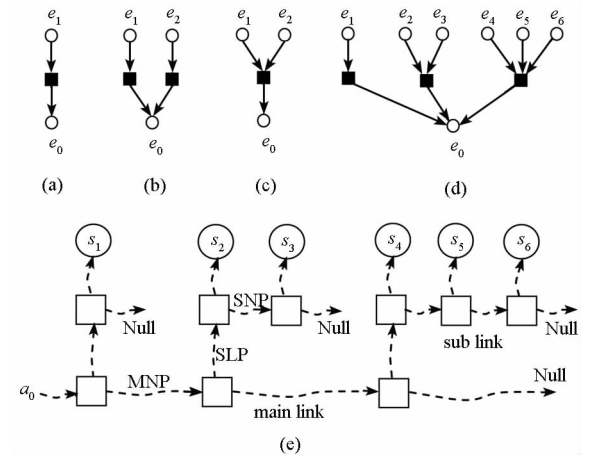


图3 警报关联值的计算

Fig. 3 Computation of alert correlation values

该结构包含2级链表,主链表 ML (main-link) 的每个控制块称为主链表控制块,对应每个节点在攻击图中的前序属性节点。每个前序属性节点对应的前序利用节点由子链表 SL (sub-link) 维护。每个子链表上的控制块包含指向该前序利用节点的状态(0 或 1)。每个主控制块指向下一主控制块的指针记为 MNP (Main-link Next Pointer),指向对应子链表的指针记为 SLP (Sub-Link

Pointer),子链表控制块中指向下一控制块的指针记为 SNP (Sub-link Next Pointer)。基于上述数据结构,警报关联度的计算方法如图4所示。

算法依次遍历主链表和子链表,并确保每个子链表中指向所有前序利用节点边的警报关联度增加和为1。若子链表中有状态为1的节点,增加值在这些状态为1的节点间均分,否则在所有节点间均分。显然,上述算法具有 $O(N)$ 的计算复杂度和 $O(N^2)$ 的存储复杂度。警报状态向量初始值全为0,其维护比较简单,此处不再赘述。

输入:关联矩阵 C ,状态向量 S ,警报 a_i
输出:关联矩阵 C
变量:temp,记录每个 sub link 中有多少个警报的状态为 TRUE

(1) 根据警报 a_i 查找相关 main link 入口,获取指向 main link 控制块的 MNP 指针;
(2) 若 MNP = Null,转到(3);否则开始执行步骤(2.1);
(2.1) 从 main link 控制块中获取 SLP 指针,temp = 0;
(2.2) 读取 SL 控制块,若相应警报状态为 TRUE,则 temp = temp + 1;
(2.3) 若 SNP = Null,转步骤(2.4),否则将 SNP 指向的 SL 控制块作为当前控制块,转到步骤(2.2);
(2.4) 根据 2.1 获得的 SLP 指针,重新遍历 sub link,若 $s_j = \text{TRUE}$,则 $c_{i,j} = c_{i,j} + \frac{1}{\text{temp}}$;
(2.5) 读取当前 ML 控制块 MNP 指向的下一个 ML 控制块作为当前 ML 控制块,转步骤(2);
(3) 算法结束

图4 警报关联度的计算方法

Fig. 4 Algorithm of alert correlation weights computation

3 网络脆弱性量化评估

图5(a)反映了一个用攻击图描述的网络攻击场景。属性 v_1 到 v_5 分别表示拥有主机1到主机5的 root 权限。其中 v_1 为初始属性, v_5 为最终攻击结果。 e_1 和 e_2 表示主机1和2分别具有脆弱性 CVE-2008-4250(服务器远程代码执行漏洞)。主机2和主机3可以相互访问且都能访问主机4,主机4可以访问攻击目标主机5。而主机4和5分别存在脆弱性 CVE-2006-3747 (Apache 缓冲区溢出漏洞)和 CVE-2004-0330 (Serv-U FTP 服务器缓冲区溢出漏洞)。

由图5(a)所示攻击图生成的警报关联图如图5(b)所示。设根据产生的网络 IDS 警报计算的每条边的警报关联度如图5(b)所示。根据上述警报关联度可以计算攻击者沿着每条攻击路径

进行攻击的情况。如表 1 所示。

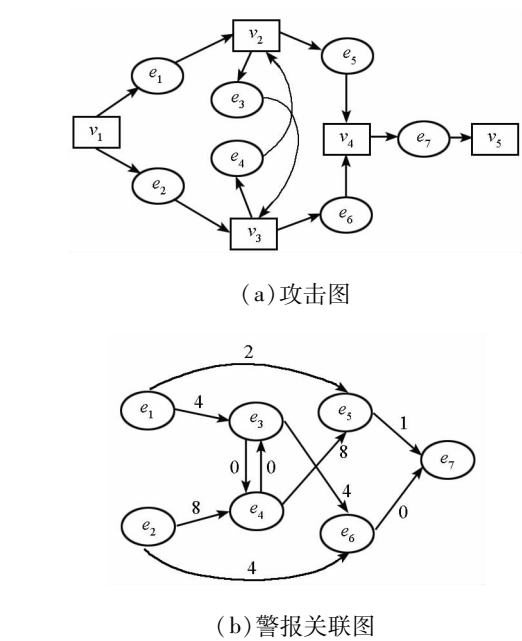


图 5 警报关联图应用示例

Fig. 5 Example of the application of alert correlation graph

表 1 攻击路径的度量值计算

Tab. 1 Computation of attack path weights

攻击路径	度量值
$e_1 - e_5 - e_7$	$2 + 1 = 3$
$e_1 - e_3 - e_6 - e_7$	$4 + 4 + 0 = 8$
$e_1 - e_3 - e_4 - e_5 - e_7$	$4 + 0 + 8 + 1 = 13$
$e_2 - e_6 - e_7$	$4 + 0 = 4$
$e_2 - e_4 - e_5 - e_7$	$8 + 8 + 1 = 17$
$e_2 - e_4 - e_3 - e_6 - e_7$	$8 + 0 + 4 + 0 = 12$

由于每条关联边的度量值计算依据 IDS 真实的警报信息,因此表 1 中攻击路径的度量值计算真正反映了攻击者的攻击行为。特别是 $e_2 - e_4$ 以及 $e_4 - e_5$ 间的关联值较大,反映了攻击者沿主机 1 到主机 3,再到主机 2 的攻击次数较多。针对这一特点,在网络防护时可采取专门的措施进行安全加固。

与文献[8]中反映 IDS 警报关联的结果图(result graph)相比,警报关联图没有记录每个警报关联的具体信息,而是根据攻击图中描述的攻击路径记录警报之间关联的次数,不但减小了计算和存储开销,而且结果更加宏观,更有利于动态分析攻击者的攻击行为。

由于关联边权值计算的复杂度为 $O(N)$,只与攻击图中利用节点的数目有关,而与 IDS 警报的数目无关,因此警报关联图可以在线计算、分析、展示并预测攻击者的攻击路径,实时反映受保护网络的安全状况,及时地根据攻击者的行为对网络进行加固。同时,根据警报关联图对 IDS 警

报的历史数据进行分析,可以获取攻击者真正关心、也是最需要防护的网络脆弱点。

4 结束语

本文提出的警报关联图概念实质上是对攻击图在时间上进行扩展,把 IDS 检测到的警报信息映射到攻击路径上,通过计算警报关联边的权值对网络脆弱性进行分析。

本文只是提出警报关联图的基本概念,更多的因素需要在下一步研究中考虑。一是目前度量值计算方法中,只考虑状态向量中记录的前序警报是否发生,而在基于攻击图的 IDS 警报推测中,每个属性节点都可能处于 TRUE、HYP 或 FALSE 状态。因此可以利用 IDS 警报对应的前序属性节点的状态进一步地对其前序警报进行区分,优化警报关联度的计算。更加重要的是 IDS 警报的缺失对警报关联计算影响很大,如何有效利用警报的预测和推断信息进行关联度计算对提高网络脆弱性分析准确度也具有重要意义。

参考文献(References)

[1] 林闯,任丰原. 可控可信可扩展的新一代互联网[J]. 软件学报, 2004,15(12):1815-1821.
LIN Chuang, REN Fengyuan. Controllable, trustworthy and scalable new generation internet[J]. Journal of Software, 2004,15(12):1815-1821. (in Chinese)

[2] Ammann P, Wijesekera D, Kaushik S. Scalable graph-based network vulnerability analysis[C]// Proceedings of the 9th ACM Conference on Computer and Communications Security, 2002:217-224.

[3] Ou X M, Boyer W F, McQueen M A. A scalable approach to attack graph generation[C]// Proceedings of the 13th ACM Conference on Computer and Communications Security, 2006: 336-345.

[4] Chen F, Su J S, Zhang Y. A scalable approach to full attack graphs generation[C]// Proceedings of the 1st International Symposium on Engineering Secure Software and Systems, 2008:150-163.

[5] Jha S, Sheyner O, Wing J. Two formal analyses of attack graphs[C]// Proceedings of 15th IEEE Computer Security Foundations Workshop, 2002:49-63.

[6] Wang L Y, Singhal A, Jajodia S. Toward measuring network security using attack graphs[C]// Proceedings of 3rd International Workshop on Quality of Protection, 2007: 49-54.

[7] Wang L Y, Islam T, Long T, et al. An attack graph-based probabilistic security metric[C]// Proceedings of 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security, 2008:283-296.

[8] Wang L Y, Liu A Y, Jajodia S. Using attack graphs for correlating, hypothesizing, and predicting intrusion alerts[J]. Computer Communications, 2006,29:2917-2933.

[9] 陈锋,张怡,苏金树,等. 攻击图两种形式化分析[J]. 软件学报, 2010,21(4):838-848.
CHEN Feng, ZHANG Yi, SU Jingshu, et al. Two formal analyses of attack graphs[J]. Journal of Software, 2010,21(4): 838-848. (in Chinese)

双预测结构融合的复杂红外图像背景抑制算法*

赵 菲, 张志勇, 肖山竹, 卢焕章
(国防科技大学 ATR 国防科技重点实验室, 湖南 长沙 410073)

摘 要: 针对复杂红外图像的背景抑制问题, 提出了一种双预测结构融合的复杂红外图像背景抑制算法。算法以图像中每个像素为中心, 在其局部区域内根据灰度相似程度和空间分布相关性计算像素的核值相似程度, 然后根据核值相似程度有侧重地融合两种不同结构的空域滤波器的输出作为最终的预测结果, 最后将原始图像与预测图像相减完成背景抑制。实验结果表明, 本文提出的算法能够较好地抑制不同红外图像中的复杂背景, 算法结构简单, 具有较强的应用价值。

关键词: 核值相似程度; 预测融合; 背景抑制; 复杂红外图像
中图分类号: TN911.73 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0113-06

Two-prediction-architecture fused complex IR images
background suppression algorithm

ZHAO Fei, ZHANG Zhiyong, XIAO Shanzhu, LU Huanzhang
(National Key Laboratory of Automatic Target Recognition (ATR),
National University of Defense Technology, Changsha 410073, China)

Abstract: Aiming at the complex background suppression problem, a two-prediction-architecture fused complex IR images background suppression algorithm is proposed. In the local region of each pixel, the Nucleus Similarity Degree (NSD) is computed on the basis of the gray value similarity and spatial distribution correlation. Then, the outputs of two spatial filters with different structures are fused as the final background prediction result. Finally, the suppression is done by subtracting the prediction from the original image. The experimental result demonstrates that the proposed algorithm can get better performance in different complex IR images while the structure of the algorithm is simple, and it is suitable for engineering application.

Key words: nucleus similarity degree; fused prediction; background suppression; complex IR image

红外凝视被动探测技术因灵敏度和帧频高、隐蔽性好、不受使用时间限制, 被广泛应用于自寻的武器系统中。一般而言, 红外探测器角分辨率有限, 当目标距离较远时, 探测器接收到的目标辐射能量较低, 同时接收到了视场内其他物体的辐射, 因此成像后的图像中大部分为空间分布不平稳的复杂背景杂波(如云背景、地面背景、海天背景等), 目标不仅信噪比较低, 而且面积较小(1个至几个像素)。

为尽量降低背景杂波的影响, 提高目标的信噪比, 在检测任务之前首先进行背景抑制。已有的背景抑制方法可以分为基于变换域^[1-3]的方法、基于时域^[4-6]的方法和基于空域^[7-10]的方法。总的来说, 基于变换域的方法是将原始图像进行傅立叶或小波变换后, 通过滤除低频的背景并保留高频的目标来达到抑制背景的目的。基于

时域的方法是以背景像素在时域上表现出较强的相关性, 而运动目标所在像素在时域上会形成尖峰为依据, 使用以前的图像对当前图像背景进行预测, 然后用原始图像减去预测图像, 完成静止背景的抑制。基于空域的方法根据背景在空间上相关性较强而目标呈现出“奇异点”的特点, 在空域内对背景进行预测, 然后用原始图像减去预测图像, 完成背景抑制。基于变换域和基于时域的方法虽然可以获得良好的背景抑制效果, 但在实际武器系统中, 成像帧频往往较高($\geq 60\text{Hz}$), 从探测器采集、传输到信息处理单元完成对图像所有的处理(目标检测、跟踪、识别等)一般只有几毫秒到十几毫秒, 对算法的实时性要求很高; 同时, 武器系统自身的姿态变化将导致帧间图像存在抖动, 像素在时域上的相关性很难保证。相比之下, 基于空域的背景抑制方法具有更好的实用价值。

* 收稿日期: 2011-06-01
基金项目: 国家部委资助项目; 国家 863 计划资助项目
作者简介: 赵菲(1983—), 男, 甘肃天水人, 博士研究生, E-mail: f_z2010@126.com;
卢焕章(通信作者), 男, 教授, 博士, 博士生导师, E-mail: luhuanzhang@nudt.edu.cn

基于空域的背景抑制算法将当前像素所在局部区域内的像素作为输入,使用线性或非线性的预测结构对图像背景进行预测,然后使用原始图像与预测得到的背景图像相减完成背景抑制。由于图像内各部分的相关性不尽相同,使用单一的预测结构很难保证在图像各处都达到较好的预测效果。例如图像中边缘与背景区域的空间相关性不同,使用相同的预测结构会使得边缘处的背景预测性能下降,导致残差图像中依然存在较多杂波,进而影响后续的检测性能。如果能够根据图像中不同的相关特性选择适合的背景预测方法,将会取得更好的效果。基于这样的想法,本文提出一种基于核值相似程度、双预测结构融合的背景抑制算法。算法以核值相似程度描述当前像素与局部区域内像素的相似性,并根据该指标融合两种不同预测结构的输出作为最终的预测结果,然后将原始图像与预测背景相减的残差作为最后的输出。算法以滑动窗口的方式处理图像中的像素,结构简单,能够以硬件加速的方式实现。

1 双预测结构融合的复杂红外图像背景抑制算法

1.1 算法原理

包含有弱小目标的复杂背景红外图像序列可描述为

$$F(x, y, k) = F_T(x, y, k) + F_B(x, y, k) + \text{Noise}(x, y, k) \quad (1)$$

其中, $F(x, y, k)$ 为红外图像的灰度值, $F_T(x, y, k)$ 为点目标灰度值, $F_B(x, y, k)$ 为背景灰度值, $\text{Noise}(x, y, k)$ 为零均值的高斯噪声, k 代表图像帧号, x 和 y 代表图像中的二维坐标。对于复杂红外背景图像而言,大面积的背景内部变化缓慢,呈现低频特性;背景的边缘因具有较剧烈的灰度变化呈现高频特性;弱小目标在图像中表现为奇异点并且呈现出高频特性。正是由于包含弱小目标的复杂背景红外图像中各种分量所表现出的这种特性,才能够使用周围像素对当前像素的背景进行预测,但由于局部区域内的相关性不同,很难利用一种预测结构获得满意的背景预测结果。

本文提出以核值相似程度为基础,融合双预测结构进行复杂红外图像的背景抑制。算法首先根据被预测像素与局部范围内像素的相关性计算被预测点的核值相似程度,然后根据核值相似程度融合两种不同预测结构的预测结果作为最后的背景预测输出,最后用原始图像减去预测图像完

成复杂背景抑制。本文算法的结构与一般背景预测算法的结构对比如图 1 所示。

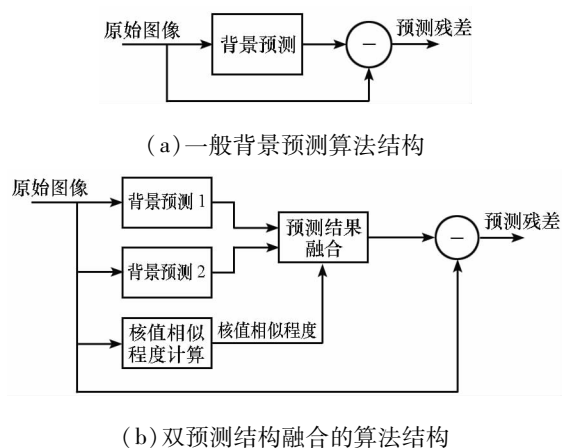


图 1 双预测结构融合算法与一般背景预测算法结构对比

Fig. 1 Architecture comparison of two predictor fused algorithm and normal spatial background prediction algorithm

1.2 核值相似程度

假设当前被预测点的位置为 (i, j) , $M \times N$ 个像素构成点 (i, j) 的局部区域,点 (i, j) 作为局部区域的中心被称为核^[11],如图 2 所示。核值相似程度是指核与 $M \times N$ 局部区域内其他像素的相似程度。针对复杂背景红外图像中弱小目标只包含灰度信息的特点,核与 $M \times N$ 窗口内像素的相似程度从两个方面考虑:

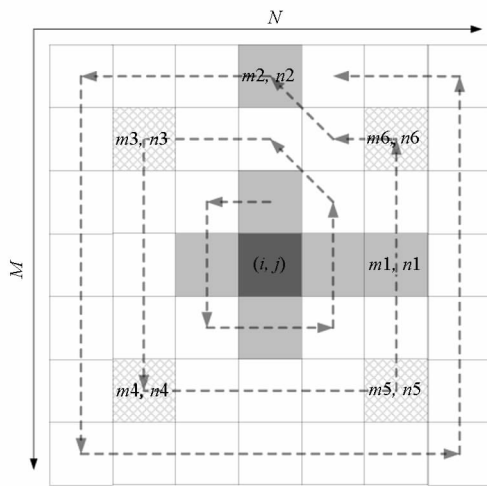


图 2 局部区域示意图

Fig. 2 Figure of local area

(1) 灰度相似程度。 $M \times N$ 窗口内,属于背景的像素往往具有数量较多的灰度相似像素;而目标在局部区域内的灰度相似像素的数量则较少。

(2) 空间分布的相关程度。在灰度相似的基础上,背景像素在空间分布上往往呈现较强的相关性,背景像素连通性较强,分布较广;而弱小目标因呈现为奇异点,局部区域内与其相连的像素

较少,分布较集中。

综合考虑灰度相似程度和空间分布的相关程度,本文算法定义点 (i,j) 的核值相似程度为

$$S(i,j) = \sum_m \sum_n \text{dis}\{(i,j), (m,n)\} \times cc(m,n) \quad (2)$$

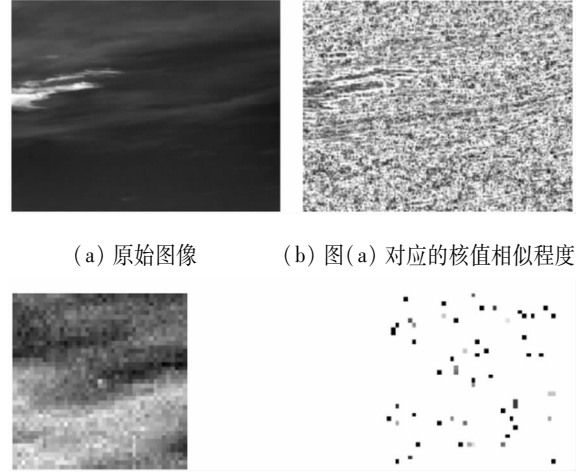
其中 $m \in [i - M/2, i + M/2], n \in [j - N/2, j + N/2], m$ 与 n 皆为整数; $\text{dis}\{(i,j), (m,n)\}$ 代表像素 (m,n) 与核之间的距离,用来衡量空间分布的广度,为方便计算,算法中采用绝对值距离,即 $\text{dis}\{(i,j), (m,n)\} = |m - i| + |n - j|$; $cc(m,n)$ 代表灰度相似像素 (m,n) 与核的连通程度,需根据像素 (m,n) 与核的灰度相似情况及空间连通情况综合判断。令 $M \times N$ 窗口内像素的访问顺序为从内而外的“包围圈”方式,如图2中虚线所示。

$cc(m,n)$ 的赋值需要在灰度相似的基础上进行判断,以 $H(m,n)$ 代表像素 (m,n) 与核 (i,j) 的灰度相似程度,定义 $H(m,n) = \begin{cases} 1, & |G(i,j) - G(m,n)| < T_g \\ 0, & |G(i,j) - G(m,n)| \geq T_g \end{cases}$,其中 $G(i,j)$ 为核的灰度值, $G(m,n)$ 为像素 (m,n) 的灰度值, T_g 为判断阈值, T_g 的取值决定着灰度相似的限度。考虑 T_g 应具有一定的鲁棒性并且计算应简单,文章中选取 $T_g = [\max(G(m,n)) - G_{\text{mean}}] \times \mu$,其中 $\max(G(m,n))$ 和 G_{mean} 分别代表局部区域内的灰度最大值和均值, μ 代表系数,一般可取0.2至0.4。

在灰度相似的基础上,算法中对 $cc(m,n)$ 赋值的准则为: $cc(m,n) = 1$,当 $H(m,n) = 1$ 并且像素 (m,n) 与核连通;否则 $cc(m,n) = 0$ 。算法中像素 (m,n) 与核连通的判据为:像素 (m,n) 与核之间至少存在一条由灰度相似的像素构成的路径。算法以由内而外的“包围圈”方式逐层判断像素的连通性,不仅判断每个像素与核的灰度相似性,同时需要判断其内圈像素中是否有像素与核有连通。以图2为例阐述 $cc(m,n)$ 的赋值准则,假设图中灰色区域代表符合灰度相似条件的像素,按照灰色虚线所示的方式扫描局部区域的所有像素,当扫描到第2圈时,点 $(m1,n1)$ 位于包围圈的右边缘,且 $cc(m1,n1 - 1) \neq 0$,因此 $cc(m1,n1) = 1$;点 $(m2,n2)$ 位于包围圈的上边缘,但 $cc(m2 + 1,n2 - 1)$ 、 $cc(m2 + 1,n2)$ 和 $cc(m2 + 1,n2 + 1)$ 均为0,所以 $cc(m2,n2) = 0$ 。位于每层4个角上的像素点,例如图中标识为阴影的点 $(m3,n3)$ 、 $(m4,n4)$ 、 $(m5,n5)$ 和 $(m6,n6)$,分别判断内层与

其相邻的像素,即 $cc(m3 + 1,n3 + 1)$ 、 $cc(m4 - 1,n4 + 1)$ 、 $cc(m5 - 1,n5 - 1)$ 和 $cc(m6 + 1,n6 - 1)$ 。从以上的定义可以看出, $cc(m,n)$ 不仅衡量像素 (m,n) 与核的灰度相似性,同时也衡量像素 (m,n) 与核的空间连通性。

根据式(2)可知,核值相似程度 $S(i,j)$ 的大小取决于局部区域内其他像素与核的连通程度以及距离。弱小目标面积小,在局部区域内的连通像素少, $S(i,j)$ 较小;背景像素因具有大面积的连通特点,因此 $S(i,j)$ 较大。本文对典型的红外云背景图像计算了其核值相似程度,其结果如图3所示。从图中可以看出,本文提出的核值相似程度能够较好地反映图像中背景和目标的差异,云层内部和边缘都具有较高的核值相似程度值,而目标的核值相似程度则较低;从图(b)和图(d)中可以看出,部分噪声也具有较低的核值相似程度值,通过背景预测结果融合后,噪声可能会被检测出来,但是噪声点在时域连续性难以保证,难以形成类似目标的运动轨迹。



(a) 原始图像

(b) 图(a)对应的核值相似程度

(c) 图(a)中目标周围原始图像 (d) 图(c)对应的核值相似程度

图3 复杂背景的核值相似程度图

Fig. 3 NSD of complex background

根据目标和背景在核值相似程度上的差异,本文融合了不同的预测结构的预测结果作为最终的预测输出。

1.3 预测结果融合

根据对图像中每个像素核值相似程度的计算,可以有效地区分出图像中存在的目标和背景像素。本文算法根据核值相似程度的不同取值,使用了两种预测结构,即有利于突出目标和有利于预测背景的两种结构,考虑算法的复杂性,两种预测结构计算都较为简单。其中有利于突出目标的预测结构具有固定的系数模板,具体为

$$\begin{bmatrix} 1/32 & 1/32 & 1/32 & 1/32 & 1/32 \\ 1/32 & 1/16 & 1/16 & 1/16 & 1/32 \\ 1/32 & 1/16 & 0 & 1/16 & 1/32 \\ 1/32 & 1/16 & 1/16 & 1/16 & 1/32 \\ 1/32 & 1/32 & 1/32 & 1/32 & 1/32 \end{bmatrix},$$

从模板的系数可以看出,该预测结构只使用周围像素对中心像素进行预测,当中心像素经过与原始图像相差后,能够突出中心像素。

本文采用灰度相似像素低通滤波结构作为有利于预测背景的预测结构,其表达式为

$$F_1(i,j) = \sum_{m=i-M/2}^{i+M/2} \sum_{n=j-N/2}^{j+N/2} G(m,n) \times \frac{H(m,n) \times \text{gaussian}(m,n)}{W} \quad (3)$$

其中 $\text{gaussian}(m,n)$ 代表以 (i,j) 为原点的二维高斯滤波器, W 是归一化系数。该预测结构只有灰度与中心像素相近的像素(即 $H(m,n) = 1$ 的像素)参与预测,能更好地预测背景像素的灰度。

假设有利于突出目标和有利于预测背景的两种结构的预测结果分别为 $F_1(i,j)$ 和 $F_2(i,j)$, 预测结果的融合表达式为 $F(i,j) = w_1 \times F_1(i,j) + w_2 \times F_2(i,j)$, 其中 w_1 和 w_2 为根据核值相似程度选定的权值,并满足 $w_1 + w_2 = 1$ 。

核值相似程度值 $S(i,j)$ 较高时,说明核所在的点 (i,j) 为背景的可能性较大,因此可以赋值较大的 w_2 和较小的 w_1 ;反之,点 (i,j) 为目标的可能性较大,则应赋予较大的 w_1 和较小的 w_2 ,同时考虑当 $S(i,j)$ 稍大时,仍然可能存在目标的可能性, w_1 和 w_2 的赋值应考虑一定的冗余。综合以上考虑,本文算法设计 w_1 和 w_2 为 $S(i,j)$ 的分段函数:

$$[w_1, w_2] = \begin{cases} [1, 0], & S(i,j) < T_{s1} \\ \left[\frac{T_{s2} - S(i,j)}{T_{s2} - T_{s1}}, \frac{S(i,j) - T_{s1}}{T_{s2} - T_{s1}} \right], & T_{s1} \leq S(i,j) \leq T_{s2} \\ [0, 1], & S(i,j) > T_{s2} \end{cases} \quad (4)$$

其中 T_{s1} 和 T_{s2} 是判断点 (i,j) 是目标或是背景相应的核值相似度阈值。本文算法中假设目标为点目标,成像后面积小于 3×3 ;由于点目标弥散后周围像素的能量较低,较难满足与目标中心像素的灰度相似性,基于这样的考虑,算法中选取 $T_{s1} = 3, T_{s2} = 9$ 。

经过 $F_1(i,j)$ 与 $F_2(i,j)$ 融合后的结果作为最后的背景预测值,通过与原图像相减得到背景抑制的结果,即残差图像。

2 实验结果与分析

本文以实际红外探测器采集的包含弱小目标

的复杂背景红外图像序列为实验样本,将本文提出的算法和基于空域的背景预测算法如 TDLMS 算法^[7]、形态学 TopHat 滤波算法^[10] 和最大化背景模型算法^[9] 等进行了比较。

为保证实验数据具有一定的代表性,本文选取的实验图像序列有以天空和云层为背景拍摄的红外图像、以地面为背景拍摄的红外图像和以海面及天空为背景拍摄的红外图像,每种类型各包括两组序列,所有序列均为 235×300 的 8 位图像,如图 4 中(a) ~ (f) 所示,图像中用方块标识的部分为目标,其中(a) 包含两个目标:T1 和 T2。



(a) 天空和云层背景图像 1



(b) 天空和云层背景图像 2



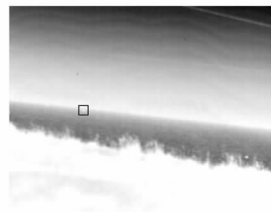
(c) 地面背景图像 1



(d) 地面背景图像 2



(e) 海面 and 天空背景图像 1



(f) 海面 and 天空背景图像 2

图 4 实验图像

Fig. 4 Experiment images

进行实验时,本文算法中关键参数的选取为 $M = N = 7$;TDLMS 算法和最大化背景模型算法的窗口大小选为 3×3 ,学习速率为 10^{-6} ,TopHat 算法的结构元素为“十”字型,边长为 3。4 种方法对图 4 所示的 6 幅图像的背景抑制结果如图 5 所示。其中图像的第 1 列是 TDLMS 算法输出的残差图像,第 2 列是形态学 TopHat 滤波输出的残差图像,第 3 列是最大化背景抑制算法输出的残差图像,第 4 列是本文算法输出的残差图像。从滤波残差的结果可以看出,本文算法的滤波残差相对其他 3 种算法明显较小,相对具有较好的背景杂波抑制效果。

本文分别统计了 6 组实验序列中所有图像的

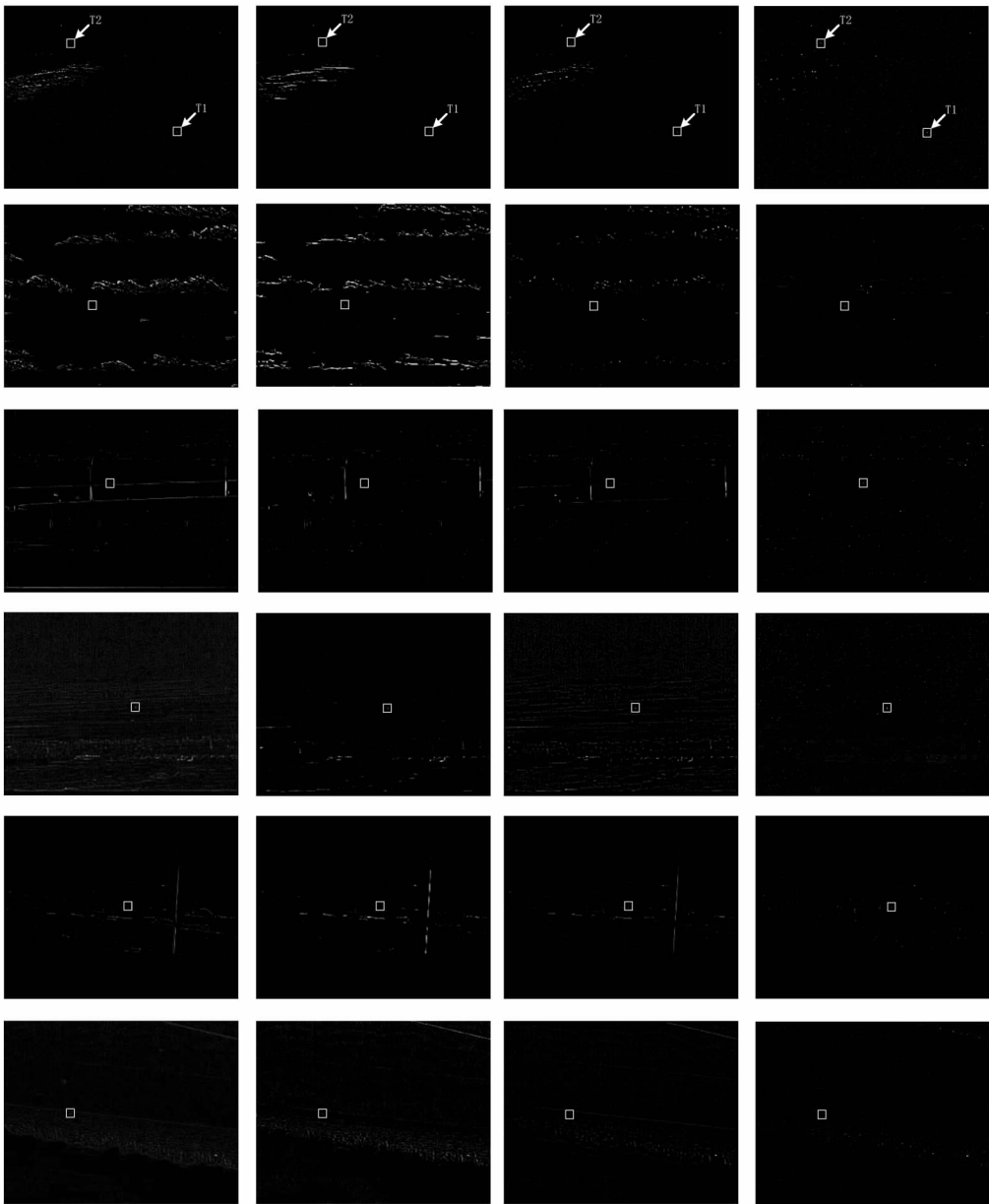


图 5 背景抑制结果
Fig. 5 Background suppression results

目标输入信噪比 SNR_{in} 和图像输入标准差 σ_{in} , 4 种背景抑制算法对应的目标信噪比增益 $Gain(SNR)$ 和标准差增益 $Gain(\sigma)$ ^[12]。其中目标信噪比定义为: $SNR = \frac{T-m}{\sigma}$, 其中 T 为目标的灰度值, m 为灰度均值, σ 为灰度标准差; 目标信噪比增益定义为: $Gain(SNR) = 20\log \frac{SNR_{out}}{SNR_{in}}$, SNR_{out} 是目标在残差图像中的信噪比; 图像标准差增益定义为: $Gain(\sigma) = 20\log \frac{\sigma_{out}}{\sigma_{in}}$, σ_{out} 是残差图像的标准差。由于原始实验图像存在背景杂波的影响, 统计目标在整幅图像中的信噪比无法客观地表现目标的强度, 因此本文在以目标为中心

的 9×9 的窗口内统计目标输入信噪比 SNR_{in} , 尽可能地排除背景的影响; 而通过背景抑制后, 图像应该表现为目标加高斯噪声的形式, 因此本文在整幅图像内统计目标的输出信噪比 SNR_{out} 。输入标准差 σ_{in} 和输出标准差 σ_{out} 均在整幅图像内进行统计。 $Gain(SNR)$ 和 $Gain(\sigma)$ 分别代表目标增强和背景抑制的程度, 对于背景抑制而言, 希望抑制后的残差图像中目标信噪比越大越好, 而背景杂波干扰越少越好, 因此越大的 $Gain(SNR)$ 值和越小的 $Gain(\sigma)$ 值代表背景抑制算法的效果越好。对于 6 组序列的实验结果统计如表 1 所示。表中列出的结果均是序列中所有图像统计结果的均值。其中图像序列 (a) 的实验结果中斜线左边和右边的值分别代表针对 T1 和 T2 得到的结果。

表 1 6 组序列实验结果

Tab. 1 Experiment results of six image sequences

图像 序列 组号	SNR_{in}	σ_{in}	算法	$Gain(SNR)$ /dB	$Gain(\sigma)$ /dB
(a)	4.26/ 2.55	39.45	TDLMS	0.06/2.32	-17.29
			TopHat	0.65/2.41	-16.93
			最大化	1.20/2.97	-17.94
			本文	10.70/12.87	-25.52
(b)	4.50	138.16	TDLMS	-6.82	-21.79
			TopHat	-5.02	-22.53
			最大化	-3.68	-25.61
			本文	5.02	-30.66
(c)	1.21	42.80	TDLMS	8.03	-21.10
			TopHat	12.90	-22.90
			最大化	10.50	-25.50
			本文	18.30	-31.90
(d)	1.33	44.50	TDLMS	9.35	-15.05
			TopHat	11.50	-24.91
			最大化	10.00	-23.70
			本文	17.30	-31.94
(e)	2.10	70.30	TDLMS	5.10	-22.00
			TopHat	7.20	-23.10
			最大化	6.84	-23.10
			本文	13.50	-28.20
(f)	1.19	66.2	TDLMS	12.40	-29.40
			TopHat	14.80	-31.10
			最大化	15.64	-32.22
			本文	23.6	-38.70

从表中对比可以看出,本文提出的算法的目标增强效果 $Gain(SNR)$ 和背景抑制效果 $Gain(\sigma)$ 明显优于其他 3 种算法。在第 2 组序列中其他 3 种算法背景抑制后的全局信噪比甚至低于原图像的局部信噪比,并未达到较好的抑制背景的目的,而本文算法对于 6 组实验序列均具有较好的背景抑制效果。这主要归因于本文算法相对其他方法能够较细致地分辨像素的局部相关特性,并能有侧重地选取更适合的背景抑制方法。

3 总结与展望

本文提出了一种双预测结构融合的复杂红外图像背景抑制算法,该算法先计算像素在局部区域内的核值相似程度,再根据核值相似程度有侧重地融合不同背景预测结果。实验证明,本文提出的算法能够较好地预测复杂红外图像中的背景分量,进而达到抑制背景和增强目标的效果。

下一步将重点研究并丰富核值相似程度的定义,融合更高效的背景抑制方法,进而取得更好的效果。同时,满足不同复杂背景情况下参数的自

适应设置也是下一步的研究内容。

参考文献(References)

[1] 杨磊,杨杰,郑忠龙. 海空复杂背景中基于自适应局部能量阈值的红外小目标检测[J]. 红外与毫米波学报, 2006, 25(1): 41-45.
YANG Lei, YANG Jie, ZHENG Zhonglong. Detecting infrared small targets based on adaptive local energy threshold under sea-sky complex backgrounds[J]. Journal of Infrared and Millimeter Waves, 2006, 25(1): 41-45. (in Chinese)

[2] 杨杰,杨磊. 基于红外背景复杂程度描述的小目标检测算法[J]. 红外与激光工程, 2006, 36(3): 382-386.
YANG Jie, YANG Lei. Small target detection algorithm based on infrared background complex degree description[J]. Infrared and Laser Engineering, 2006, 36(3): 382-386. (in Chinese)

[3] 王文龙,韩保君,张红萍. 一种海空背景下红外小目标检测新算法[J]. 光子学报, 2009, 38(3): 725-728.
WANG Wenlong, HAN Baojun, ZHANG Hongping. A new algorithm of small target detection for infrared image in background of sea and sky[J]. Acta Photonica Sinica, 2009, 38(3): 725-728. (in Chinese)

[4] 徐剑峰,吴一全,周建江. 基于时域背景预测检测红外图像序列中的小目标[J]. 中国图像图形学报, 2007, 12(9): 1598-1603.
XU Jianfeng, WU Yiquan, ZHOU Jianjiang. Small target detection based on temporal predictions of background in infrared image sequences[J]. Journal of Image and Graphics, 2007, 12(9): 1598-1603. (in Chinese)

[5] 王博,刘德连,张建奇. 基于背景移除的时域目标检测[J]. 通信学报, 2009, 30(7): 67-72.
WANG Bo, LIU Delian, ZHANG Jianqi. Temporal filtering for target detection algorithm based on static background elimination[J]. Journal on Communications, 2009, 30(7): 67-72. (in Chinese)

[6] 武斌,姬红兵,李鹏. 基于三阶累积量的红外弱小运动目标检测新方法[J]. 红外与毫米波学报, 2006, 25(5): 364-367.
WU Bin, JI Hongbing, LI Peng. New method for moving dim target detection based on third order cumulate in infrared image[J]. Journal of Infrared and Millimeter Waves, 2006, 25(5): 364-367. (in Chinese)

[7] Tarun S, James R Z. Performance evaluation of 2-d adaptive prediction filters for detection of small objects in image data[J]. IEEE Transaction on Image Processing, 1993, 2(3): 327-340.

[8] 朱红,赵亦工. 基于背景自适应预测的红外弱小运动目标检测[J]. 红外与毫米波学报, 1999, 18(4): 305-310.
ZHU Hong, ZHAO Yigong. Detection of weak and small moving infrared targets by adaptive prediction of background[J]. Journal of Infrared and Millimeter Waves, 1999, 18(4): 305-310. (in Chinese)

[9] 徐军,向健华,梁昌洪. 最大化背景模型用于检测红外图像中的弱小目标[J]. 光子学报, 2002, 31(12): 1483-1486.
XU Jun, XIANG Jianhua, LIANG Changhong. Small target detection based on maximum background model in IR images[J]. Acta Photonica Sinica, 2002, 31(12): 1483-1486. (in Chinese)

[10] Zhang F, Li C F, Shi L. Detecting and tracking dim moving point target in IR image sequence[J]. Infrared Physics and Technology, 2005, 46: 323-328.

[11] Smith S M, Brady J M. SUSAN-a new approach to low level image processing[J]. International Journal of Computer Vision, 1997, 23(1): 45-78.

[12] Tartakovsky A, Blazek R. Effective adaptive spatial temporal technique for clutter rejection inIRST[C]// SPIE Proceedings: Signal and Data Processing of Small Targets. Orlando, FL, 2000: 85-95.

单极化圆台相控阵天线的空域极化特性*

何 密,李永祯,肖顺平
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘 要: 针对水平或垂直极化偶极子激励的单极化圆台相控阵天线,建立了其电扫描时空域极化特性的数学模型。理论分析和仿真结果表明单极化圆台相控阵天线的空域极化比随观测角偏离法线方向的角度单调递增,随斜面倾角增大而单调递减。同时,圆台相控阵天线的方向图与阵元极化方式、波束指向角以及观测角有关。圆台面规模越大,方向图和空域极化变化越丰富。以上结论为基于圆台相控阵天线空域极化特性的极化散射矩阵测量、极化滤波和抗干扰等研究提供了理论依据,为全极化圆台相控阵天线设计、极化特性分析和极化校准等研究奠定了理论基础。

关键词: 空域极化特性;圆台阵;相控阵;共形阵;
中图分类号: TP958 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0119-06

Spatial polarization characteristics of the single-polarization truncated cone phased array antenna

HE Mi, LI Yongzhen, XIAO Shunping
(college of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: For a single-polarization-truncated-cone-phased array antenna stimulated by horizontal or vertical polarized dipoles, the spatial polarization characteristics are modeled when the array antenna electronically scans in space. The theoretical analysis and simulations show that: the spatial polarization ratio of a single-polarization-truncated-cone-phased array monotonically increases as the angles between the pointing angles or observation angles and the normal increases; the spatial polarization ratio monotonically decreases as the slope angles decrease; the patterns of a single-polarization-truncated-cone-phased array are related to the polarization of array elements, the pointing angles, and the observation angles; the larger the size of a single-polarization-truncated-cone-phased array antenna is, the more abundant the spatial polarization information is. These conclusions offer a theoretical foundation for studies such as the polarization scattering matrix measurement, the polarization filtering, and the polarimetric anti-jamming based on the spatial polarization characteristics of the single-polarization-truncated-cone-phased array. The conclusions offer a theoretical basis for the studies such as the antenna design, the polarization characteristics analysis and the polarization calibration of the full-polarization-truncated-cone array as well.

Key words: spatial polarization characteristics; truncated cone array; phased-array; conformal array

近年来,对各类天线空域极化特性的研究越来越引起人们的关注^[1-2]。研究单极化天线空域极化特性能够在不增加设备量的前提下有效提高雷达系统的极化测量、跟踪、识别和抗干扰的能力,具有重要的理论和现实意义。文献[2-5]中提出了基于单极化天线空域极化特性的目标极化散射矩阵测量、抗有源干扰和真假目标鉴别等算法。文献[1-5]中主要分析了机械扫描雷达的空域极化特性,文献[6-8]拓展到对二元阵和平面相控天线空域极化特性的分析。很多情形下,雷达需要附形于雷达载体,以减小空气阻力或雷达散射截面积^[9-10]。目前,共形阵列天线的空域

极化特性鲜见报道,相对于平面阵列,共形阵列复杂的几何结构给空域极化特性的研究带来困难,考虑到天线空域极化特性的研究具有重要的现实意义,因此,有必要研究共形阵列天线的空域极化特性。

关于旋转轴对称的共形阵,具有方位向对称的优点,通过轮换激励阵元的一列,实现360°覆盖扫描并且保持方位向方向图不变^[9-11]。圆台阵列天线是一种常见的关于旋转轴对称的共形阵,通常共形在弹头和机翼等雷达载体上。为了研究单极化的圆台相控阵列天线的空域极化特性,首先建立圆台相控阵天线空域极化特性的模

* 收稿日期:2011-09-01
基金项目:国家自然科学基金重点资助项目(60736006);国防科技大学优秀研究生创新资助项目(B090405)
作者简介:何密(1982—),女,四川营山人,博士,E-mail: hmcherry@126.com;
肖顺平(通信作者),男,教授,博士,博士生导师,E-mail: xiaosp@gmail.com

型,然后分析了水平、垂直极化偶极子激励时圆台阵列天线在不同的观测角和波束指向角下的空域极化比,并且给出了斜面倾角、波束指向角对极化比和天线增益等的影响。

1 圆台相控阵天线空域极化特性的建模

圆台阵列天线的结构示意图如图 1 所示。设圆台顶的半径为 R_0 ,高为 H ,斜边倾角为 α ,圆台的中心位于全局坐标系的原点 O 。圆台阵列由 $M \times N$ 个阵元构成,其中 M 为行数, N 为列数。设圆台阵列天线的纵向阵元间隔在 z 轴上的投影长度为 d_1 ,第 m 层阵元在 z 轴上坐标为 z_m ,则圆台第 m 层圆周的半径可表示为

$$R_m = R_0 + \left(\frac{H}{2} - z_m \right) \cot(\alpha) \quad (1)$$

第 m 层圆周上阵元间隔为

$$d_2(m) = \frac{2\pi R_m}{N} \quad (2)$$

圆台阵列天线在阵面中部可同时形成多个波束,从而 360° 扫描覆盖整个方位向空域。考虑到圆台阵列天线的有效孔径、遮挡效应以及系统复杂度^[9-10],同时波束通常为 3 到 4 个。为更好地分析圆台阵列天线的空域极化特性,下面给出四分之一圆台阵列形成一个波束时的相关公式推导,其他波束的推导类似。设四分之一圆台阵面关于矢量 \mathbf{AB} 对称, A 点是 x 轴与四分之一圆台阵面的交点, \mathbf{AB} 位于 xOz 平面内且与 x 轴的夹角为 $\frac{\pi}{2} - \alpha$,矢量 \mathbf{AB} 垂直于圆台阵面,是四分之一圆台阵面的法线方向,具有方向角 ($\theta = \alpha, \varphi = 0$)。圆台阵列交换激励阵元的一列可实现方位向的扫描并保持天线方位向的极化方向图不变,此时的方位角分辨率为 $\frac{2\pi}{N}$,若需要更高的方位角分辨率,每列阵元可在方位向上做小范围相描。

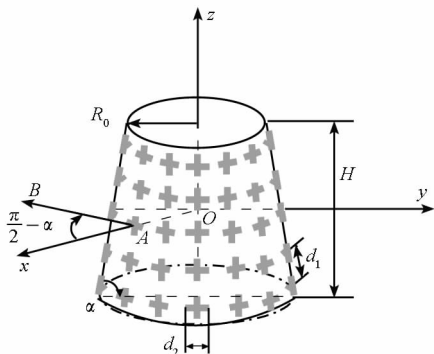


图 1 圆台相控阵天线的结构示意图

Fig. 1 The sketch map of a truncated cone phased-array antenna

为了推导圆台阵列天线的空域极化特性,需要建立 3 个坐标系,即全局坐标系、旋转坐标系和局部坐标系,三者的示意图见图 2。全局坐标系 $O-xyz$ 由单位矢量 \mathbf{a}_x 、 \mathbf{a}_y 和 \mathbf{a}_z 组成,全局坐标系中某点的坐标可由笛卡尔坐标 (x, y, z) 表示,也可由球坐标 (θ, φ, r) 表示。为方便描述圆台阵面上的点,需建立旋转坐标系。旋转坐标系是由单位矢量 $\mathbf{a}_{x'}$ 、 $\mathbf{a}_{y'}$ 和 \mathbf{a}_z 组成,它是通过将全局坐标系中的 x 轴、 y 轴绕 z 轴旋转 φ_n 构成,其中 φ_n 为第 n 列阵元的方位角。设极化方向图观测点与原点 O 的距离为 r 、方向为 (θ, φ) ,其局部坐标系由单位矢量 \mathbf{a}_φ 、 \mathbf{a}_θ 和 \mathbf{a}_r 组成,分别为 φ 、 θ 增加的方向和矢量 \mathbf{r} 的方向。旋转坐标系和局部坐标系的关系为

$$\begin{bmatrix} \mathbf{a}_\theta \\ \mathbf{a}_\varphi \\ \mathbf{a}_r \end{bmatrix} = \begin{bmatrix} \cos\theta\cos(\varphi - \varphi_n) & \cos\theta\sin(\varphi - \varphi_n) & -\sin\theta \\ -\sin(\varphi - \varphi_n) & \cos(\varphi - \varphi_n) & 0 \\ \sin\theta\cos(\varphi - \varphi_n) & \sin\theta\sin(\varphi - \varphi_n) & \cos\theta \end{bmatrix} \begin{bmatrix} \mathbf{a}_{x'} \\ \mathbf{a}_{y'} \\ \mathbf{a}_z \end{bmatrix} \quad (3)$$

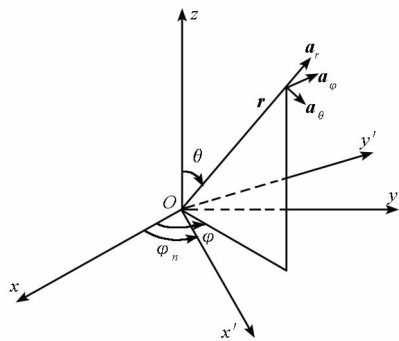


图 2 3 个坐标系关系

Fig. 2 Three coordinate systems

圆台阵列表面的每个阵元可由水平极化偶极子(h)或者垂直极化偶极子(v)组成。水平极化偶极子沿圆周的切线方向(即 \mathbf{a}_φ 方向)放置,取方位角 φ 增加的方向为正;垂直极化偶极子沿斜边方向放置,取沿斜边向上的方向为正。由于圆台阵面共形的特点,垂直极化偶极子的激励电流同时具有 \mathbf{a}_z 方向和 $\mathbf{a}_{x'}$ 方向的分量。第 mn 个阵元偶极子 q ($q = x', y'$ 或 z) 的电场为^[12]

$$\mathbf{E}_{mn}^q(r) = -\frac{k^2 e^{-jkr}}{4\pi\epsilon r} \mathbf{a}_r \times \mathbf{a}_r \times \mathbf{M}_{mn}^q \quad (4)$$

其中 $k = 2\pi/\lambda$ 为波数, λ 为波长, r 为观察点距原点的距离, $\epsilon = 8.85 \times 10^{-12} \text{ Fm}^{-1}$ 为自由空间的介电常数, $I_{mn} = a_{mn} e^{j\beta_{mn}}$ 是第 mn 个阵元的激励电流,幅度为 a_{mn} ,相位为 β_{mn} , \mathbf{M}_{mn}^q 为偶极子 q 电流的激励方向矢量。令 $\mathbf{M}_{mn}^{q'} \triangleq \mathbf{a}_r \times \mathbf{a}_r \times \mathbf{M}_{mn}^q$,当 $\mathbf{M}_{mn}^q = \mathbf{a}_{x'}$, $\mathbf{M}_{mn}^q = \mathbf{a}_{y'}$ 和 $\mathbf{M}_{mn}^q = \mathbf{a}_z$ 时,分别有

$$\begin{aligned} \mathbf{M}'_{mn} &= \mathbf{a}_{x'} [1 - \sin^2 \theta \cos^2 (\varphi - \varphi_n)] - \\ &[\mathbf{a}_{y'} \sin \theta \sin (\varphi - \varphi_n) + \mathbf{a}_z \cos \theta] \sin \theta \cos (\varphi - \varphi_n) \end{aligned} \quad (5.1)$$

$$\begin{aligned} \mathbf{M}'_{mn} &= \mathbf{a}_{y'} [1 - \sin^2 \theta \sin^2 (\varphi - \varphi_n)] - \\ &[\mathbf{a}_{x'} \sin \theta \cos (\varphi - \varphi_n) + \mathbf{a}_z \cos \theta] \sin \theta \sin (\varphi - \varphi_n) \end{aligned} \quad (5.2)$$

$$\begin{aligned} \mathbf{M}'_{mn} &= \mathbf{a}_z \sin^2 \theta - \\ &[\mathbf{a}_{x'} \cos (\varphi - \varphi_n) + \mathbf{a}_{y'} \sin (\varphi - \varphi_n)] \sin \theta \cos \theta \end{aligned} \quad (5.3)$$

观测点到圆台阵面上第 mn 个阵元与到原点的相位差为

$$k(z_m \cos \theta + R_m \sin \theta \cos (\varphi - \varphi_n)) \quad (6)$$

若波束指向角为 (θ_0, φ_0) , 则第 mn 个阵元所需的相移为

$$-k(z_m \cos \theta_0 + R_m \sin \theta_0 \cos (\varphi_0 - \varphi_n)) \quad (7)$$

假设每个阵元都是理想阵元, 仅均匀向外辐射, 并且阵元间的耦合已被消除。若记

$$\begin{aligned} A_{mn} &= -\frac{k^2 e^{-jkr} I_{mn}}{4\pi\epsilon r} \cdot \\ &e^{jk[z_m(\cos \theta - \cos \theta_0) + R_m[\sin \theta \cos (\varphi - \varphi_n) - \sin \theta_0 \cos (\varphi_0 - \varphi_n)]]} \end{aligned} \quad (8)$$

则阵元均为水平极化偶极子(记为 h)的圆台相控阵天线的辐射场为

$$\mathbf{E}^h = \sum_{mn} A_{mn} \mathbf{M}'_{mn} \quad (9)$$

阵元均为垂直极化偶极子(记为 v)的圆台相控阵天线的辐射场为

$$\mathbf{E}^v = \sum_{mn} A_{mn} (\mathbf{M}'_{mn} \cos \alpha + \mathbf{M}'_{mn}^z \sin \alpha) \quad (10)$$

天线远场的共极化分量和交叉极化分量常采用 Ludwig 定义^[11]。局部水平、垂直极化基 \mathbf{e}_h 和 \mathbf{e}_v 分别为^[13]

$$\begin{cases} \mathbf{e}_h = \mathbf{a}_\varphi \\ \mathbf{e}_v = -\mathbf{a}_\theta \end{cases} \quad (11)$$

第 mn 个阵元为水平极化偶极子的方向矢量在 \mathbf{e}_h 和 \mathbf{e}_v 上的投影矢量为

$$\begin{bmatrix} \mathbf{M}'_{mn} \cdot \mathbf{e}_h \\ \mathbf{M}'_{mn} \cdot \mathbf{e}_v \end{bmatrix} = \begin{bmatrix} \cos (\varphi - \varphi_n) \\ -\sin (\varphi - \varphi_n) \cos \theta \end{bmatrix} \quad (12)$$

第 mn 个阵元为垂直极化偶极子的方向矢量在 \mathbf{e}_h 和 \mathbf{e}_v 上的投影矢量为

$$\begin{bmatrix} \mathbf{M}'_{mn}^v \cdot \mathbf{e}_h \\ \mathbf{M}'_{mn}^v \cdot \mathbf{e}_v \end{bmatrix} = \begin{bmatrix} -\sin (\varphi - \varphi_n) \cos \alpha \\ \cos (\varphi - \varphi_n) \sin \theta \cos \theta \cos \alpha + \sin \theta \sin \alpha \end{bmatrix} \quad (13)$$

阵元均为 p 极化偶极子的圆台相控阵天线的共极化(记为 co-pol) 和交叉极化(记为 x-pol) 电场分量为

$$\begin{bmatrix} E_{\text{co-pol}}^p \\ E_{\text{x-pol}}^p \end{bmatrix} = \sum_{mn} A_{mn} \begin{bmatrix} \mathbf{M}'_{mn} \cdot \mathbf{e}_p \\ \mathbf{M}'_{mn}^p \cdot \mathbf{e}_o \end{bmatrix}, p \neq o \quad (14)$$

空域极化比定义为

$$R_a^p(\theta, \varphi) = \frac{E_{\text{co-pol}}^p}{E_{\text{x-pol}}^p} \quad (15)$$

从式(8)、(14)、(15)可以看出, 单极化圆台相控阵天线空域极化比与波束指向角 (θ_0, φ_0) 、观测角 (θ, φ) 、阵面行列数、阵元间隔、斜边倾角以及激励电流等有关。当观测角 $\theta = 90^\circ$ 时, 由式(12)和(14)可以推出 $E_{\text{x-pol}}^h = 0$, 从而极化比为 $-\infty$ 。由于圆台阵列阵元的方向图与单元的位置有关, 使得阵因子无法提取出来, 从而式(14)和(15)没有简洁的表达式, 但可从公式上粗略推导出单极化圆台相控阵天线空域极化比关于波束指向角、观测角、阵面行列数以及斜边倾角等变量的变化关系。从式(1)可推出, R_m 关于斜面倾角 α 单调递减, 对于阵元为水平极化偶极子的情形, 天线空域极化比是斜面倾角的单调减函数; 对于阵元为垂直极化偶极子的情形, 主极化分量关于 α 的偏导数比上交叉极化分量关于 α 的偏导数约为 $-\cot(\alpha)$, 在 α 的取值范围 $[0^\circ, 90^\circ]$, $-\cot(\alpha) \leq 0$, 因此, 空域极化比也关于 α 单调递减。对单极化圆台相控阵天线的共极化和交叉极化分量而言, 由于阵元的加权因子 A_{mn} 是一致的, 粗略估计时只考虑投影矢量关于俯仰和方位观测角的偏导数, 可以推出当俯仰观测角偏离 90° 时, 空域极化比单调递增, 当方位观测角偏离 0° 时, 空域极化比单调递增。单极化圆台相控阵天线方向图的有效口径随观察角偏离波束指向方向而降低, 从而使得天线增益降低。圆台阵列天线的规模会导致求和结果式(14)的变化, 如果 m 和 n 越大, 求和结果的细节变化越丰富。

此外, 阵元间隔会影响天线方向图的形状, 阵元间隔过大会导致天线方向图出现栅瓣, 文献[9]中给出了线阵和平面阵出现栅瓣的条件, 然而对于共形阵的栅瓣条件没有显式表达式^[11], 文中暂不考虑阵元间隔的影响, 统一设置较小的阵元间隔。

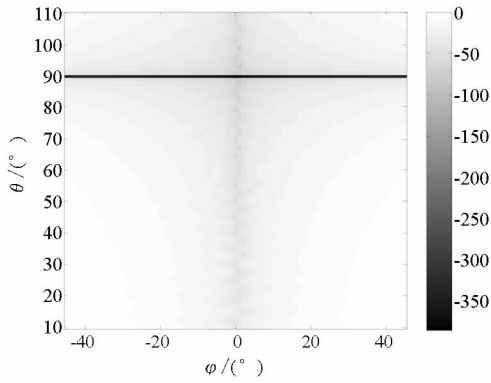
2 仿真分析

假设圆台相控阵天线的每个阵元等幅激励, 波长 λ 为 0.1m, 纵向阵元间隔在 z 轴上的投影长度为 $\frac{\lambda}{2}$, 仿真条件为: 斜面倾角分别为 40° 、 50° 、 60° 、 70° 、 80° 和 90° , 波束指向在法线方向和偏离法线方向, 阵列大小分别为 11×11 、 17×17 以及

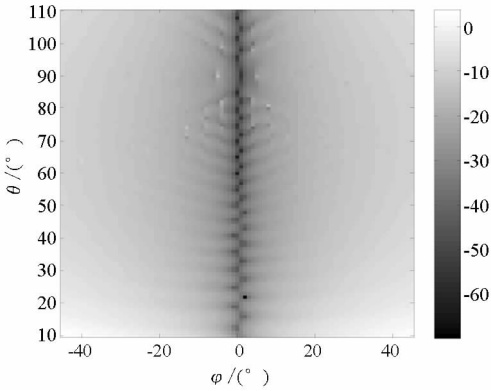
21 × 21。仿真中不同波束指向角以及俯仰和方位观测角下水平、垂直极化偶极子激励圆台阵的方向图,均以波束指向角和俯仰观测角在法线方向上时水平极化偶极子激励的方向图主瓣峰值进行归一化。

2.1 空域极化比随俯仰和方位观测角的变化具有一定规律

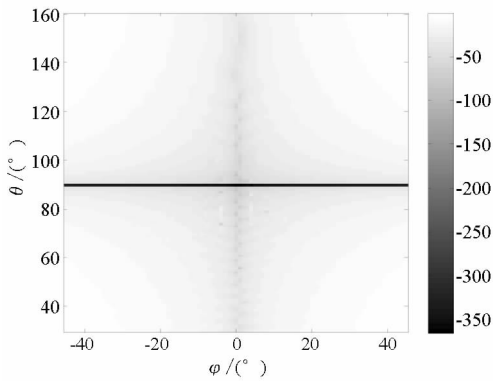
波束指向角无论在法线方向还是偏离法线方向时,空域极化比随俯仰和方位观测角的变化规律一致:水平极化偶极子激励时,俯仰向上空域极化比在 $\theta = 90^\circ$ 时最小,随着天顶角偏离 $\theta = 90^\circ$ 而增大,方位向上 $\varphi = 0^\circ$ 附近时空域极化比最小,随着方位角偏离 $\varphi = 0^\circ$,极化比增大;垂直极化偶极子激励时,当方位角大于方位向主瓣宽度一半时,空域极化比随着天顶角偏离波束指向角而增大,方位向上 $\varphi = 0^\circ$ 附近空域极化比最小,随着方位角偏离 $\varphi = 0^\circ$ 空域极化比增大。以 21 × 21 规模的圆台阵、斜面倾角为 60° 、波束指向角在法线方向和 $(100^\circ, 0^\circ)$ 为例,如图 3 所示,可以看出空域极化比的变化规律如上所述。



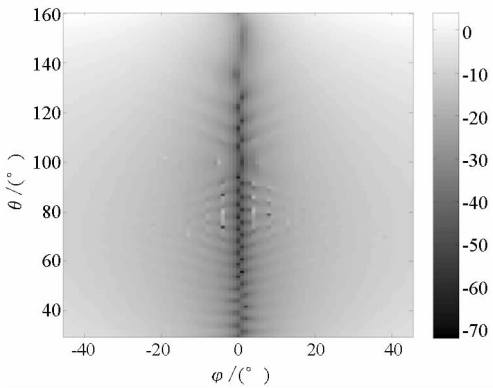
(a) 水平极化偶极子激励且波束指向在法线方向



(b) 垂直极化偶极子激励且波束指向在法线方向



(c) 水平极化偶极子激励且波束指向角为 $(100^\circ, 0^\circ)$

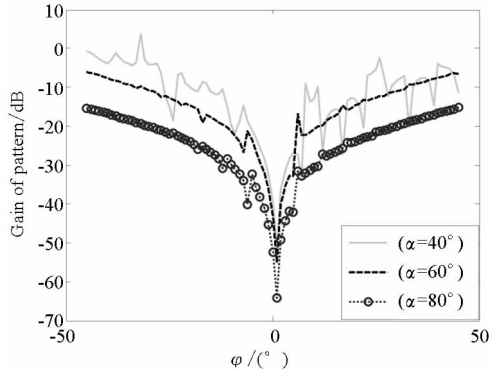


(d) 垂直极化偶极子激励且波束指向角为 $(100^\circ, 0^\circ)$

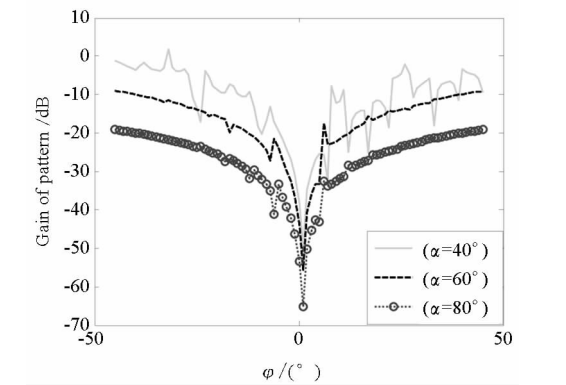
图 3 21 × 21 圆台阵的空域极化比
Fig. 3 The spatial polarization ratio of a quarter of truncated cone array antenna with 21 × 21 elements

2.2 空域极化比随斜面倾角变化的关系

空域极化比随着斜面倾角的增大而减小,极端的情形是柱形阵和锥形阵,柱形阵的空域极化比小于锥形阵,圆台阵的空域极化比大小介于两者之间。以 21 × 21 规模的圆台阵为例,当斜面倾角变化时,法线方向上空域极化比的变化规律如图 4 所示。



(a) 水平极化偶极子激励



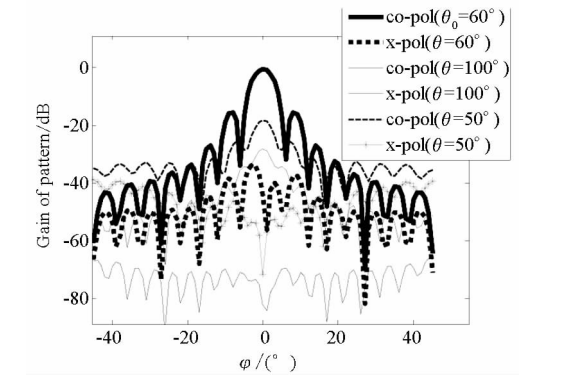
(b)垂直极化偶极子激励

图 4 不同斜面倾角下波束指向在法线方向时的空域极化比

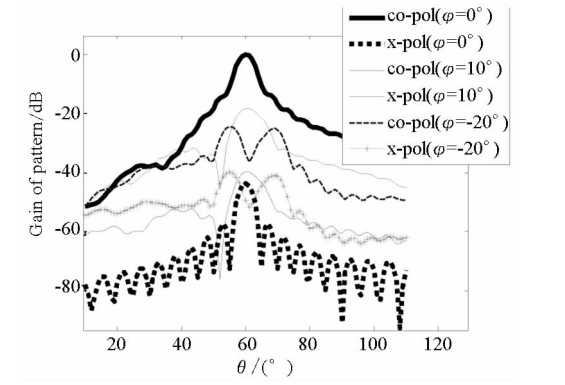
Fig.4 The spatial polarization ratio with different slope angles when the beam points along the normal direction

2.3 方向图与俯仰和方位观测角的关系

水平与垂直极化偶极子激励的主极化方向图随着俯仰观测角偏离波束指向方向,波束的增益降低,远离 $\varphi=0^\circ$ 的方位向电平严重抬升,波束指向角偏离法线方向越大,越容易导致波纹效应产生^[11]。在圆台阵列天线的主瓣内,水平、垂直极化偶极子激励的主极化方向图随着方位观测角偏离波束指向方向,天线的增益降低,远离 $\varphi=0^\circ$ 的方位向电平严重抬升,当方位观测角大于圆台阵列天线的主瓣宽度一半时,由于圆台阵列的非平面几何结构将导致方向图出现波纹效应,如果减小阵元间隔或者增大斜面倾角,波纹效应会减轻^[11]。以斜面倾角为 60° 的 21×21 规模的圆台阵为例,如图 5 和图 6 所示,当波束指向分别在法线方向和 $(100^\circ, 0^\circ)$,俯仰和方位观测角变化时,空域极化比的变化规律如上所述,图 5(b) 中 $\varphi=-20^\circ$ 时,方向图出现波纹效应,图 6(b) 中 $\varphi=10^\circ$ 时出现波纹效应。

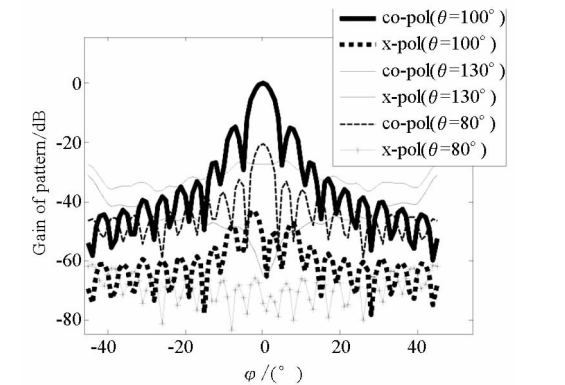


(a)水平极化偶极子激励不同俯仰角上

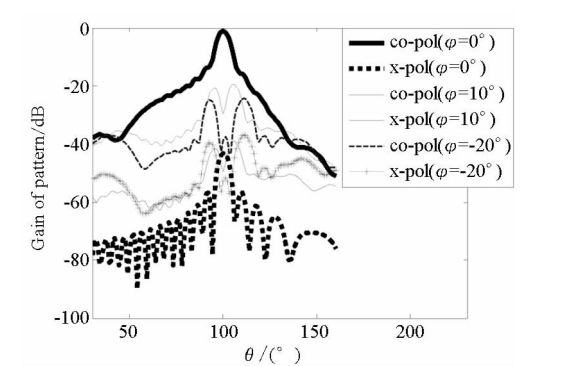


(b)垂直极化偶极子激励不同方位角上

图 5 21×21 圆台阵波束指向在法线方向的方向图
Fig.5 The patterns of a quarter of truncated cone array antenna with 21×21 elements when the beam points along the normal direction



(a)水平极化偶极子激励不同俯仰角上



(b)垂直极化偶极子激励不同方位角上

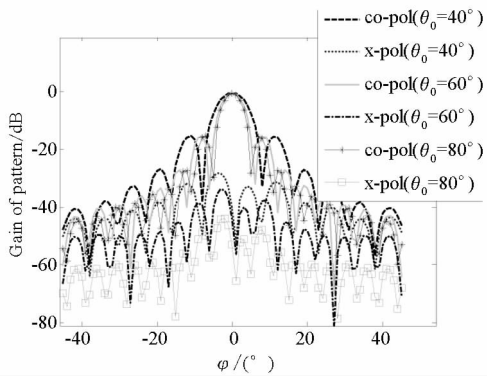
图 6 波束指向角为 $(100^\circ, 0^\circ)$ 时 21×21 圆台阵在不同方位角上的方向图

Fig.6 The patterns of a quarter of truncated cone array antenna with 21×21 elements when the beam points along the direction $(100^\circ, 0^\circ)$

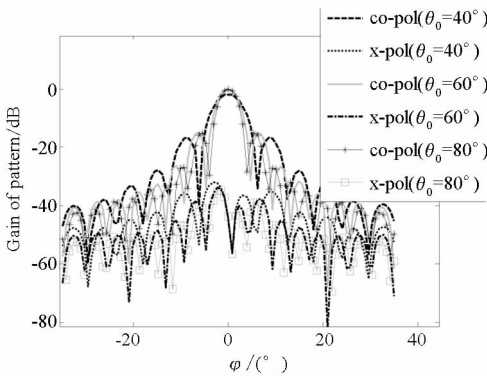
2.4 方向图与波束指向角的关系

对于水平极化偶极子激励的情形,主瓣最大增益不随波束指向角的变化而变化,当波束指向角(天顶角)增大时,主瓣变窄,旁瓣降低。对于垂直极化偶极子激励的情形,当波束指向偏离法

线方向时天线的增益降低,当波束指向角(天顶角)增大时,主瓣变窄,旁瓣升高。以斜面倾角为 60° 的 21×21 规模的圆台阵为例,图 7 给出在不同的波束指向角(天顶角)上,水平与垂直极化偶极子分别激励的共极化和交叉极化方向图。



(a) 水平极化偶极子



(b) 垂直极化偶极子

图 7 不同波束指向角上 21×21 圆台阵激励的方向图

Fig. 7 The patterns with different pointing angles of a quarter of truncated cone array antenna with 21×21 elements

2.5 方向图、空域极化比与阵面规模的关系

取相同的阵元间隔时,阵面规模越大,方向图的主瓣越窄,而且所包含的旁瓣信息越丰富。相对应的空域极化比,阵面规模越大,所含的信息量越丰富。以阵元规模为 11×11 的圆台阵为例,斜面倾角为 60° ,波束指向在法线方向上,其他仿真参数不变,图 8 给出垂直极化偶极子激励的空域极化比,对比图 3(b)可以看出,图 8 中的空域极化比正好是图 3(b)中间部分信息的放大。

3 结 论

本文主要分析了单极化圆台相控阵天线这一较常用的共形阵列天线的空域极化特性,圆台相控阵天线的方向图跟阵元极化方式、波束指向角

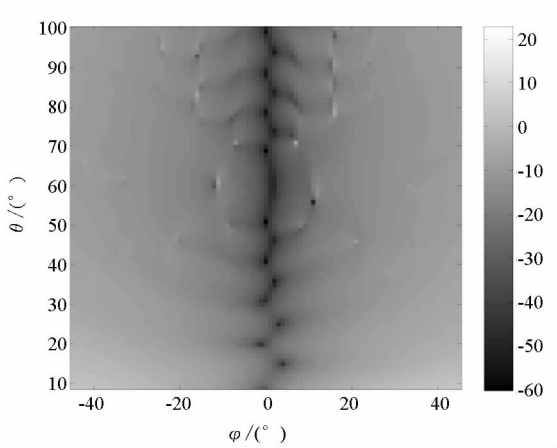


图 8 规模为 11×11 的圆台阵列天线空域极化比
Fig. 8 The spatial polarization ratio of a quarter of truncated cone array antenna with 11×11 elements

以及观测角有关。空域极化比随着斜面倾角的增大而降低。水平极化偶极子激励时,俯仰向上空域极化比在天顶角 $\theta = 90^\circ$ 时最小,空域极化比随着观测角偏离法线方向而增大;垂直极化偶极子激励时,空域极化比随着观测角偏离法线方向而增大。圆台阵面规模越大,空域极化比信息越丰富。

本文的研究成果对进一步利用单极化圆台阵列天线空域极化特性进行散射矩阵测量、极化滤波和抗干扰等研究具有一定理论指导意义,同时对全极化圆台相控阵天线的设计、极化特性分析和极化校准也奠定了一定理论基础。今后可以考虑圆台阵元间的耦合效应以及阵列锥削、加权等因素的影响,实际中诸如共形阵复杂的馈电网络以及阵列表面涂覆介质等内容都需要进一步深入研究。

参考文献 (References)

[1] 罗佳,王雪松,李永祯,等. 天线空域极化特性的表征及分析[J]. 电波科学学报,2008, 23(4): 620-628.
LUO Jia, WANG Xuesong, LI Yongzhen, et al. Spatial polarization characteristics for antenna[J]. Chinese Journal of Radio Science, 2008, 23(4): 620-628. (in Chinese)
[2] 罗佳. 天线空域极化特性及应用[D]. 长沙:国防科技大学,2008.
LUO Jia. Application and analysis of spatial polarization characteristics for antenna[D]. Changsha: National University of Defense Technology, 2008. (in Chinese)

STBC cooperation scheme for the multi-antenna relay [J]. Journal of Xidian University, 2010, 37(2): 334 – 339. (in Chinese)

[3] Laneman J, Wornell G. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks [J]. IEEE Transactions Inform Theory, 2003, 49(10): 2415 – 2425.

[4] Dharmawansa P, McKay M R, Mallik R K. Dual hop MIMO relaying with orthogonal space-time block codes [C]// IEEE International Conference on Communication 2009, Dresden, June 2009:1 – 5.

[5] He J, Kam P Y. Adaptive cooperative space-time block coding with amplify-and-forward strategy [C]// IEEE Vehicular Technology Conference Spring 2008, Singapore, May 2008: 1236 – 1240.

[6] He J, Kam P Y. Cooperative space-time block coding with amplify-and-forward strategy: exact bit error probability and adaptive forwarding schemes [J]. Physical Communication, 2008, 1(3): 209 – 220.

[7] Jang J, Lee K B, Lee Y H. Transmit power and bit allocations for OFDM systems in a fading channel [C]// IEEE GLOBECOM 2003, Dec 2003: 858 – 862.

[8] Li Y, Wang W B, Kong J, et al. Power allocation and subcarrier pairing in OFDM-based relaying system [C]// IEEE International Conference on Communication 2008, May 2008: 2602 – 2606.

[9] Shen Z H, Wang X X, Zhang H T. Power allocation and subcarrier pairing for OFDM based AF cooperative diversity systems [C]// IEEE Vehicular Technology Conference Spring 2009, Barcelona, April 2009: 1 – 5.

[10] Hammerstrom I, Wittneben A. On the optimal power allocation for Nonregenerative OFDM relay links [C]// IEEE International Conference on Communication 2006. Istanbul, June 2006: 4463 – 4468.

[11] Wang L J, Zhao H T, Tang Y, et al. The joint relay selection with power allocation for outage-optimization in OFDM-based cooperative relaying [C]// IEEE ICCASM 2010, Tai Yuan, China, Oct 2010: 557 – 561.

(上接第 124 页)

[3] 李金梁, 罗佳, 常宇亮, 等. 基于天线空域极化特性的虚拟极化接收技术[J]. 电波科学学报, 2009, 24(3): 389 – 393.

LI Jinliang, LUO Jia, CHANG Yuliang, et al. Virtual polarization receiver based on the spatial polarization characteristics of antenna [J]. Chinese Journal of Radio Science, 2009, 24(3): 389 – 393. (in Chinese)

[4] 罗佳, 王雪松, 李永祯, 等. 雷达目标极化散射矩阵测量的新方法研究[J]. 信号处理, 2009, 25(6): 868 – 873.

LUO Jia, WANG Xuesong, LI Yongzhen, et al. A novel polarimetric scattering matrix measurement method [J]. Signal Processing, 2009, 25(6): 868 – 873. (in Chinese)

[5] 戴幻尧, 李永祯, 王雪松, 等. 基于和差波束极化特性的目标极化散射矩阵测量方法研究[J]. 电子与信息学报, 2010, 32(4): 6 – 12.

DAI Huanyao, LI Yongzhen, WANG Xuesong, et al. A new target scattering matrix measurement algorithm based on polarization characteristics of sum-and-difference beam [J]. Journal of Electronics & Information Technology, 2010, 32(4): 6 – 12. (in Chinese)

[6] Dai H Y, Wang X S, Li Y Z, et al. Spatial polarization characteristics and scattering matrix measurement of orthogonal polarization binary array radar[J]. Science in China (Series F), 2010, 53(12): 2687 – 2695.

[7] 刘勇, 李永祯, 王雪松, 等. 有源压制干扰背景下的雷达目标空域极化检测方法[J]. 电波科学学报, 2010, 25(3): 445 – 450.

LIU Yong, LI Yongzhen, WANG Xuesong, et al. Radar target spatial polarization detection under active suppressive jamming background[J]. Chinese Journal of Radio Science, 2010, 25(3): 445 – 450. (in Chinese)

[8] 戴幻尧, 李永祯, 陈志杰, 等. 电扫偶极子相控阵天线的空域极化特性分析[J]. 国防科技大学学报, 2010, 32(1): 84 – 89.

DAI Huanyao, LI Yongzhen, CHEN Zhijie, et al. Spatial polarization characteristics of electronically scanning dipole phased arrays antenna [J]. Journal of National University of Defence Technology, 2010, 32(1): 84 – 89. (in Chinese)

[9] 张光义. 相控阵雷达系统[M]. 北京: 国防工业出版社, 1994.

ZHANG Guangyi. Phased-array radar system [M]. Beijing: Defense Industry Press, 1994. (in Chinese)

[10] Robert J M. Phased array antenna handbook [M]. 2nd ed. Artech House, 2005.

[11] Josefsson L, Persson P. Conformal array antenna theory and design[M]. IEEE Press, 2006.

[12] Ishimaru A. Wave propagation and scattering in random media [M]. IEEE Press, 1997.

[13] Ludwig A C. The definition of cross polarization[J]. IEEE Tran. on AP, 1973: 116 – 119.

一种高分辨 SAR 层析成像方法^{*}

孙希龙,余安喜,董臻,孙造宇,梁甸农
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘要:在城区建筑三维重建中,如何利用非均匀的少量观测数据,在保持方位向-距离向分辨率的同时实现高度向的高分辨成像是 SAR 层析面临的一个主要难题。在确定性点目标模型下,基于 RELAX 算法,提出了一种适用于城区建筑三维重建的 SAR 层析高分辨成像方法。与统计模型下的空间谱估计方法相比,该方法无需多视处理,能够保持方位向-距离向分辨率。与奇异值分解方法相比,该方法具有更优的高度向分辨能力。在仿真数据和 Envisat-ASAR 实测数据处理实验中,该方法的有效性得到了验证。

关键词:合成孔径雷达;层析;高分辨;松弛算法

中图分类号:TN957 **文献标志码:**A **文章编号:**1001-2486(2012)03-0125-06

A high resolution imaging method for SAR tomography

SUN Xilong, YU Anxi, DONG Zhen, SUN Zaoyu, LIANG Diannong

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: In the three-dimensional reconstruction of urban areas and man-made infrastructure, the urgent problem of SAR tomography is how to achieve satisfactory elevation resolution with a small number of non-uniform passes and preserve azimuth-range resolution simultaneously. An imaging method for SAR tomography based on RELAX algorithm is proposed in the framework of deterministic models. Compared with the spatial spectral estimation method that works statistically, this method does not need to deal with a spatial multilooking and maintains high resolution of direction dimension-distance dimension. Compared with the singular value decomposition method, this method has much better resolution of elevation dimension. Results obtained by processing simulated data and real data of ENVISAT-ASAR verified the promising potentiality of this imaging method.

Key words: Synthetic Aperture Radar; tomography; high resolution; RELAX

20 世纪末,层析成像技术被引入到合成孔径雷达(Synthetic Aperture Radar, SAR)中,产生了 SAR 层析技术(SAR Tomography, TomoSAR)^[1],该技术利用在不同空间位置获得的多次观测将传统 SAR 的方位向一维合成孔径扩展成方位-高度平面内的二维合成孔径,克服了传统 SAR 二维成像中高度向分辨能力丧失的难题,真正实现了对观测目标的三维成像,近年来受到了国内外众多研究机构和研究者的关注。

当前,SAR 层析数据通常由单天线 SAR 系统的重轨任务获得。由于当前在轨运行的单天线 SAR 系统在设计之初并未考虑 SAR 层析的应用需求,所以对同一目标的重轨观测次数较少,并且轨道间的间隔非均匀。在这种情况下,基于傅里叶变换的方法难以得到令人满意的高度向聚焦效果,旁瓣高且分辨率低。以 Capon 和 MUSIC 为代

表的基于空间谱估计的 SAR 层析成像方法^[2-3]虽然可以得到较好的高度向分辨能力,但是此类方法需要使用独立同分布的多视数据估计观测的协方差矩阵,从而降低了方位向-距离向分辨率。城区建筑的三维重建是 SAR 层析技术的重要应用领域之一^[4]。对于城区建筑目标,强散射元的尺寸小、密度高,需要独立同分布多视数据的 SAR 层析成像方法并不适用^[5]。城区建筑三维重建中,如何利用非均匀的少量观测数据,在保持方位向-距离向分辨率的同时实现高度向的高分辨成像是 SAR 层析中亟待解决的难题。

基于奇异值分解的成像方法(Singular Value Decomposition, SVD)^[6]虽然能够保持方位向-距离向分辨率,但是仅能通过压缩目标的高度向成像范围得到有限的超分辨能力。RELAX 算法^[7]是在 CLEAN 算法基础上发展起来的一种基于非

^{*} 收稿日期:2011-09-17

基金项目:国家自然科学基金资助项目(60903206,61101187);国防科技大学科研计划项目(JC10-04-03)

作者简介:孙希龙(1982—),男,山东高密人,博士研究生,E-mail:longlongsun@sohu.com;

余安喜(通信作者),男,副教授,博士,硕士生导师,E-mail:yu_anxi@sina.com

线性最小方差 (NLS) 准则的谱估计算法, 具有较好的分辨能力和鲁棒性。文献[8—9]在统计模型框架下, 提出基于 M-RELAX (Multilook RELAX) 算法的 SAR 层析成像方法, 并使用仿真实验对该方法进行了验证。但是该方法与其他空间谱估计方法相同, 都需要多视处理估计观测的协方差矩阵, 对城区建筑的 SAR 层析三维成像而言并不适用。本文基于经典 RELAX 算法^[7], 在确定性点目标模型下, 提出了适用于城区建筑三维重建的 SAR 层析成像方法。该方法对多轨单视数据进行处理, 在保持方位向-距离向分辨率的同时实现了高度向的高分辨成像。在仿真数据和 Envisat-ASAR 实测数据处理实验中, 本文方法的性能得到了很好的验证。

1 SAR 层析基本原理

为了实现对目标的三维成像, SAR 层析需要沿斜距垂向 (normal-slant-range, nsr) 形成合成孔径, 当前这主要通过单天线 SAR 系统的重轨任务模式来实现。假设对同一目标从不同的空间位置进行了 N 次观测, 经过方位向-距离向二维压缩后可得 N 幅 SAR 单视复图像。首先选取其中一幅图像作为主图像, 在通过图像配准和去斜操作后, 图像序列中某一分辨单元的复序列可表示为^[6]

$$g(n) = \int_{-s_{\max}}^{s_{\max}} \gamma(s) \exp(j2\pi\xi_n s) ds \quad (1)$$

$$\xi_n = \frac{2b_n}{\lambda r}, \quad n = 1, 2, \dots, N \quad (2)$$

其中, s 为 nsr 向高度, $[-s_{\max}, s_{\max}]$ 为目标的 nsr 向跨度, $\gamma(s)$ 为目标的复散射系数沿 nsr 向的分布函数, λ 为雷达波长, r 为主图像中心斜距, b_n 为垂直基线^[10]。

当垂直基线均匀采样、使用傅里叶变换方法进行高度向聚焦时, 根据奈奎斯特定理, 无模糊重构范围 $2s_{\max}$ 为

$$2s_{\max} = \frac{\lambda r}{2\Delta b} \quad (3)$$

其中, Δb 为垂直基线采样间隔。SAR 层析的 nsr 向等效合成孔径由垂直基线跨度决定, 与等效合成孔径对应的 nsr 向瑞利分辨率 (Rayleigh resolution) ρ_s 为

$$\rho_s = \frac{\lambda r}{2B} \quad (4)$$

其中, B 为垂直基线跨度。在单天线 SAR 系统重轨任务模式中, 对同一目标的重轨观测数日通常较少, 垂直基线跨度小并且采样非均匀。此时, 无模

糊重构范围可由平均采样间隔 $\Delta b_{ave} = B/(N-1)$ 通过 (3) 式近似得到。由 (4) 式知, 为了得到高度向的高分辨需要具有足够大的垂直基线跨度, 但是非均匀采样情况下, 基于傅里叶变换的成像方法甚至无法达到 (4) 式所示的分辨率, 所以必须寻求更优的成像方法。

2 基于 RELAX 的 SAR 层析成像

城区场景电磁散射主要包含如下几种主要成分: (1) 倾斜屋角等倾斜面造成的单次散射; (2) 屋脊和屋檐处的边缘散射; (3) 垂直面与平行面间构成的二面角造成的二次散射; (4) 复杂结构造成的三次散射; (5) 粗糙表面造成的漫散射。前 4 种散射的强度较强, 散射元尺寸小, 通常可以用点目标模型表示^[5]。由粗糙表面引起的漫散射的散射强度较弱, 通常被作为杂波。假设某一方位向-距离向分辨单元内存在 K 个强散射元, 则 (1) 式可记为如下形式,

$$g(n) = \sum_{k=1}^K \gamma(s_k) \exp(j2\pi\xi_n s_k) + e_n \quad (5)$$

$n = 1, 2, 3, \dots, N$

其中, s_k 和 $\gamma(s_k)$ 分别为第 k 个强散射元的 nsr 向高度和复散射强度。在单天线 SAR 系统重轨任务所获得的数据中, 不可避免地存在着时间去相关、大气相位、形变相位以及杂波等不利因素, 在进行 nsr 向聚焦之前必须对其进行补偿, 提高数据质量^[6,11]。(5) 式中的 e_n 表示对各种不利因素进行补偿之后的残余噪声。由于强散射元的位置 s_k 和散射强度 $\gamma(s_k)$ 是我们所关注的变量, 所以 SAR 层析的高度向聚焦可以等效为复正弦信号的参数估计问题, 待估参数为 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1}^K$ 。

RELAX 算法是在 CLEAN 算法基础上发展起来的一种基于最小非线性方差 (NLS) 准则的谱估计算法。RELAX 算法具有较好的超分辨能力, 并且对噪声不作任何限定性假设, 即使对色噪声也具有较好的鲁棒性和有效性^[7]。将 (5) 式写为矩阵形式为

$$\mathbf{g} = \mathbf{A}\boldsymbol{\gamma} + \mathbf{e} \quad (6)$$

其中

$$\mathbf{g} = [g(1) \ g(2) \ \dots \ g(N)]^T \quad (7)$$

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_K]$$

$$\mathbf{a}_k = [\exp(j2\pi\xi_1 s_k) \ \exp(j2\pi\xi_2 s_k) \ \dots \ \exp(j2\pi\xi_N s_k)]^T \quad (8)$$

$$\boldsymbol{\gamma} = [\gamma(s_1) \ \gamma(s_2) \ \dots \ \gamma(s_K)]^T \quad (9)$$

$$\mathbf{e} = [e_1 \ e_2 \ \dots \ e_N]^T \quad (10)$$

RELAX 算法通过最小化如下的 NLS 代价函数对各散射元的幅度与高度参量进行估计:

$$C_1(\{\gamma(s_k), s_k\}_{k=1}^K) = \|g - A\gamma\|^2 \quad (11)$$

式中 $\|\cdot\|$ 代表欧几里德范数。观察(11)式可知,该 NLS 代价函数对 s_k 而言是非线性函数,但对 γ 而言却是线性函数。首先将 NLS 代价函数依据 γ 最小化求得 γ 的估计为

$$\hat{\gamma} = (A^H A)^{-1} A^H g \quad (12)$$

其中 $(\cdot)^H$ 表示共轭转置操作。散射元高度 $s = [s_1 \ s_2 \ \dots \ s_K]^T$ 可通过最小化如下的代价函数获得:

$$C_2(\{s_k\}_{k=1}^K) = \|[I - A(A^H A)^{-1} A^H]g\|^2 \quad (13)$$

其中, I 为单位矩阵。求解(13)式是一个非线性多维最优化问题,需要进行全局极值的多维搜索,计算量相当大,而且可能无法求出最优解。RELAX 算法基于信号分离机理,用迭代的方法将极为复杂的多维搜索问题简化为较简单的多个一维搜索问题进行求解。

假设除第 k 个强散射元外的其他强散射元的参数都已经成功估计出,则通过(12)、(13)式得到第 k 个强散射元的参数估计为

$$\begin{aligned} \hat{\gamma}(s_k) &= \frac{(a_k)^H g_k}{(a_k)^H a_k} \Big|_{s_k = \hat{s}_k} \\ &= \sum_{n=1}^N g_k(n) \exp(-j2\pi \xi_n \hat{s}_k) / N \\ \hat{s}_k &= \arg \max_{s_k} \left\{ \left| \sum_{m=1}^M g_k(n) \exp(-j2\pi \xi_m s_k) \right|^2 \right\} \end{aligned} \quad (14)$$

其中 g_k 的定义为

$$\begin{aligned} g_k &= [g_k(1) \ g_k(2) \ \dots \ g_k(N)]^T \\ g_k(n) &= g(n) - \sum_{i=1, i \neq k}^K \gamma(s_i) \exp(j2\pi \xi_n s_i) \end{aligned} \quad (16)$$

下面给出基于 RELAX 的 SAR 层析成像的具体迭代步骤:

步骤 1 假设 $K = 1$,

通过(14)和(15)式求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1}$;

步骤 2 假设 $K = 2$,

将求得的 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1}$ 代入(16)式计算得到 g_2 , 进而通过(14)和(15)式求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=2}$;

将求得的 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=2}$ 代入(16)式计算得到 g_1 , 进而通过(14)和(15)式重新求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1}$;

重复步骤 2 直至满足内部收敛条件。具体的内部收敛条件将在后面给出。

步骤 3 假设 $K = 3$,

将求得的 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1,2}$ 代入(16)式计算得到 g_3 , 进而通过(14)和(15)式求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=3}$;

将求得的 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=2,3}$ 代入(16)式计算得到 g_1 , 进而通过(14)和(15)式重新求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1}$;

将求得的 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=1,3}$ 代入(16)式计算得到 g_2 , 进而通过(14)和(15)式重新求得 $\{\hat{\gamma}(s_k), \hat{s}_k\}_{k=2}$;

重复步骤 3 直至满足内部收敛条件。

步骤 4 重复上述相同的步骤直至 K 等于实际的强散射元个数或者预先设定的强散射元个数。

上述每一步骤中,当邻近两次迭代中由(11)式所示 NLS 代价函数的变化值小于某个门限 ε_{NLS} ,则认为达到了内部收敛。上述迭代算法中的另一个问题是强散射元个数 K 的估计问题,具体方法可参见文献^[12-13]。对于城区建筑目标,同一方向-距离向分辨单元内包含的强散射元个数由建筑的结构复杂度决定,通常为 1 至 3 个^[5]。文献[7]证明当设定的强散射元个数大于真值时,RELAX 算法依然能够得到各复正弦参量的稳健估计。所以,在实测数据处理中无需对散射元个数进行精确的估计,通常设定 $K = 3$ 就可以得到较好的聚焦效果,并且不会增加太多的额外计算负担。

3 仿真与实测数据处理

本节将利用仿真实验和实测数据处理实验验证本文方法的性能。使用的实测数据为 Envisat 自 2003 年 10 月至 2007 年 11 月间获得的 20 轨 ASAR 数据,成像目标选定为美国的安那罕天使球场 (Angel Stadium of Anaheim, 北纬 33°48'0.11", 西经 117°52'58.85")。图 1 为球场的 SAR 图像。选取 2005 年 6 月 15 日获得的图像作为主图像,主图像中心斜距 r 为 843130m,雷达波长 λ 为 0.056m,入

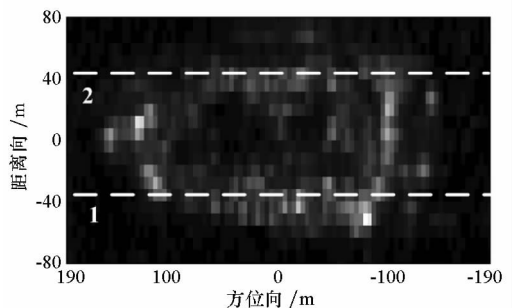


图 1 球场的 SAR 图像

Fig. 1 SAR image of Angel Stadium

射角约为 21° , 垂直基线总跨度为 1403m, 垂直基线平均采样间隔为 73.8m。由 (3) 式得无模糊重构范围 $2s_{\max}$ 约为 320m; 由 (4) 式得 nsr 向瑞利分辨率为 16.8m, 与之对应的垂直于地面的绝对高度分辨率为 6.02m。图 2 为球场的三维模型, 该球场顶棚的最大高度为 42m, 距离向远端露天看台的高度为 24m。

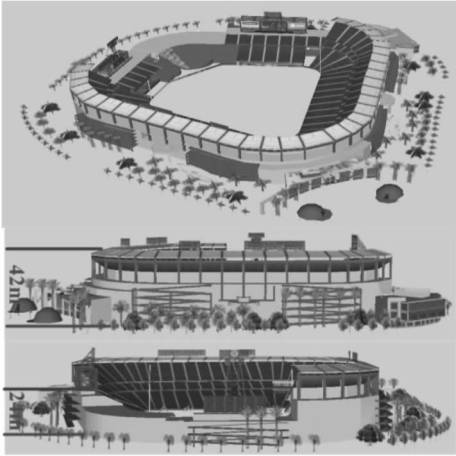


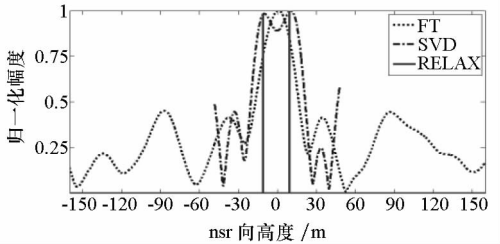
图 2 球场的三维模型

Fig. 2 Three-dimensional model of Angel Stadium

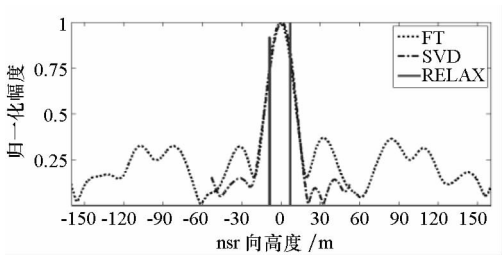
3.1 仿真实验

利用实测数据的系统和轨道参数进行仿真实验。假设加性噪声 e 的每个分量都服从零均值复高斯分布 $CN(0, \sigma^2)$ 且相互独立。首先设定 nsr 向相距 20m 的两个强散射元。两散射元的信噪比 $SNR_1 = SNR_2 = 10\text{dB}$, 其中第 i 个强散射元的信噪比定义为 $SNR_i = \frac{|\gamma_i|^2}{\sigma^2}$ 。图 3 为傅里叶变换方

法、SVD 方法与本文方法的处理结果。虽然两散射元的距离略大于 ρ_s , 但是由于非均匀基线采样的影响, 基于傅里叶变换的成像方法仍无法将其分辨, 并且结果中存在较高的旁瓣。基于奇异值分解的成像方法通过将成像范围压缩到 $0.6s_{\max}$, 两个强散射元得到了分辨。在基于 RELAX 的成像中, 设定内部收敛条件 ε_{NLS} 为 10^{-3} , nsr 向搜索步长为 1m, 最终两目标得到了成功分辨。接下来, 将两个强散射元的距离缩短到 15m (低于 ρ_s), 此时傅里



(a) 相距 20m



(b) 相距 15m

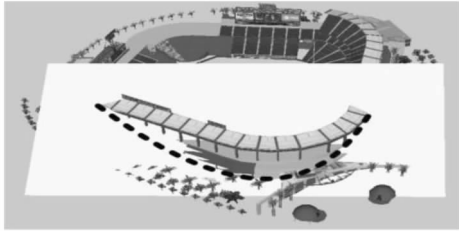
图 3 两个强散射元的成像结果

Fig. 3 Focusing results of two strong scatterers

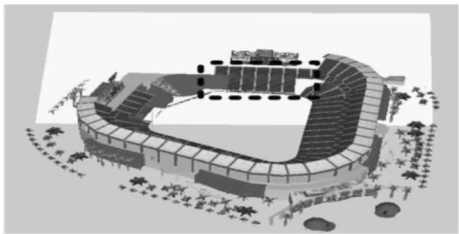
叶变换方法和 SVD 方法都无法将两个强散射元分辨开来, 而本文方法仍得到了较好的分辨效果。上面的实验证明本文方法具有明显优于 SVD 方法的分辨能力。

3.2 实测数据处理

利用 Envisat 实测数据对安那罕天使球场进行层析成像, 在成像之前使用文献 [11] 提出的方法对大气相位等进行了补偿。选取具有代表性的两个剖面进行处理, 如图 1 中的虚线所示。为了便于对成像结果进行评估, 图 4 给出了利用球场的三维模型得到的两个剖面的形状: 第一个剖面切过球场的看台顶棚和前墙体, 由于球场的金属顶棚和墙体的复杂结构会对电磁波产生较强的散射, 所以在图中的抛物线型虚线上会存在强散射元, SAR 层析成像结果也应基本呈虚线所示的形状。此外, 一些低矮附属建筑也可能带来强的散射。第二个剖面切过球场的远端看台, 由于密集排布的座椅会对电磁波产生二次或多次散射, 所以在虚线所示的矩形区域内应存在强散射元。



(a) 剖面 1



(b) 剖面 2

图 4 三维模型得到的剖面形状

Fig. 4 Sections from three dimension structure

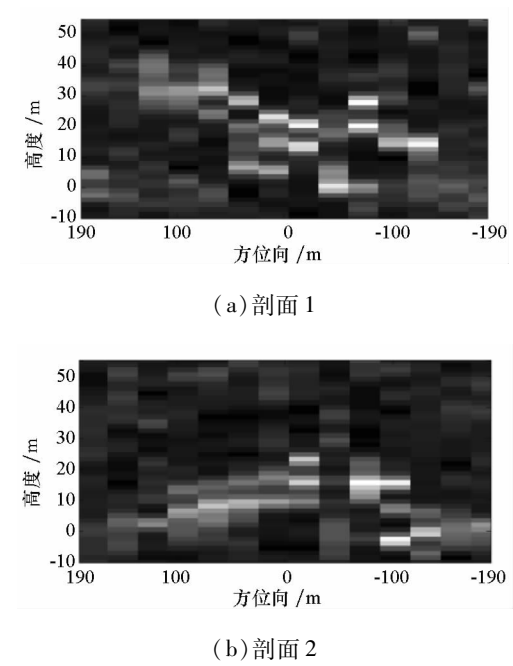


图5 Capon方法处理结果

Fig. 5 SAR Tomography results of Capon

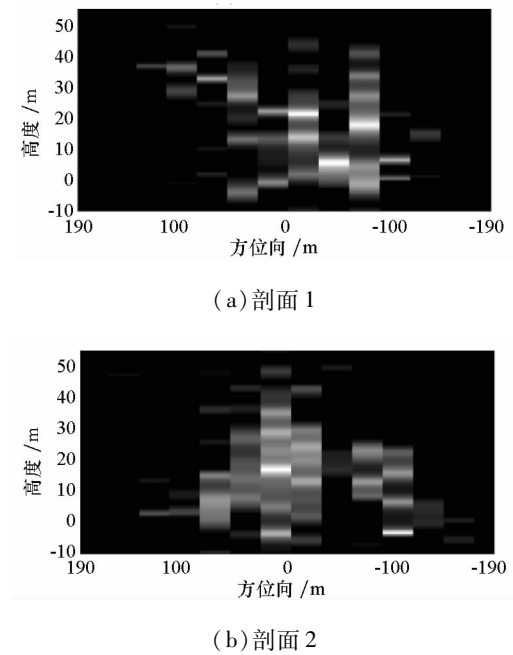
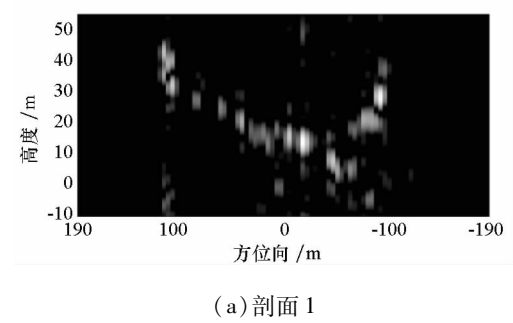


图6 MUSIC方法处理结果

Fig. 6 SAR Tomography results of MUSIC



(b) 剖面 2

图7 SVD方法处理结果

Fig. 7 SAR Tomography results of SVD

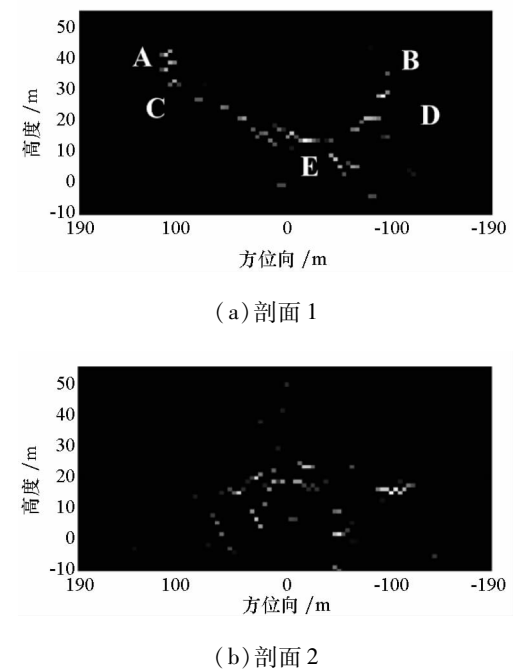


图8 RELAX方法处理结果

Fig. 8 SAR Tomography results of RELAX

首先,使用Capon和MUSIC两种方法进行成像,处理中使用 7×3 (方位向 \times 距离向)的窗口估计协方差矩阵。图5和图6分别给出了处理结果,其中nsr向高度已根据雷达入射角转换为垂直于地面的绝对高度。该两种方法处理结果的方位向分辨率明显降低且成像效果较差,基本无法反映球场的高度向结构。这主要是因为强散射元密度高、尺寸小,相邻分辨单元内存在着不同的强散射元,观测数据的数学模型无法很好地满足Capon和MUSIC方法的前提假设,相邻分辨单元的数据无法准确地等效于快拍数据。接下来,利用SVD和本文方法进行成像。在SVD方法的处理中,成像范围被压缩到 $1.2s_{\max}$,此范围已经接近于球场的实际nsr向跨度。在本文方法的处理中,设定nsr向的搜索步长为 3.4m ,迭代中内部收敛条件 ε_{NLS} 为 10^{-3} ,散射元个数 K 为3。图7和图8分别给出了处理结果。这两种方法的处理结果与图4中由球场三维模型预测的结果吻合较

好,检测到的顶棚高度为 43m,露天看台的高度为 24m,符合该球场的实际高度。虽然 SVD 方法的成像结果也较好地反映了球场的高度向结构,但是从图 7 和图 8 的对比中可见本文方法具有更优的高度向分辨能力,而且不存在 SVD 方法中的高旁瓣问题。

下面结合球场的三维模型对图 8 中散射元的散射机理进行分析。由于球场的结构较为复杂,对各散射元的散射机理进行精确分析的难度较大,所以下面的分析较为粗略。根据图 4,第一个剖面的 A 和 B 处应该存在由球场顶棚造成的散射,C 和 D 处应该存在由顶棚的四棱形支柱造成的强散射。但是,B 处未出现强散射元,C 和 D 处散射元的强度也存在较大差别。这是由于顶棚和四棱柱的散射具有较强的各向异性,因为各处的入射角不同,所以散射强度也存在较大的差异,甚至无法产生强散射。E 处的散射元应该来自于球场门廊水平屋顶与球场垂直墙面的二次散射,所以具有较强的散射强度。第二个剖面切过球场的远端看台,较强的散射元应该由座椅引起的二次散射引起,较弱的散射元应该由三次散射或者更复杂的散射机理引起。

4 结 论

基于 RELAX 算法,提出了一种高分辨 SAR 层析成像方法。该方法对多轨单视数据进行处理,能够在保持方位向-距离向分辨率的同时实现高度向的高分辨成像。在利用实测数据系统参数和轨道参数进行的仿真实验中,本文方法得到了明显优于 SVD 方法的高度向分辨率。在利用 Envisat-ASAR 实测数据对安那罕天使球场进行的 SAR 层析处理中,本文方法同样得到了较好的成像效果。本文方法对城区建筑目标三维重建的有效性得到了很好的验证。

参考文献 (References)

- [1] She Z, Gray D A, Bogner R E. Three-dimensional SAR imaging via multiple pass processing[C]//Processing of IEEE International Geoscience and Remote Sensing Symposium, 1999: 35-38.
- [2] Guillaso S, Reigber A. Scatterer characterisation using polarimetric SAR tomography [C]// Processing of IEEE International Geoscience and Remote Sensing Symposium, 2005: 2685-2688.
- [3] Lombardini F, Pardini M, Fornaro G, et al. Linear and adaptive spaceborne three dimensional SAR tomography: a comparison on real data [J]. IET Radar Sonar Navigation, 2009, 3(4): 424-436.
- [4] 张红,江凯,王超,等. SAR 层析技术的研究与应用[J]. 遥感技术与应用,2010, 25(3): 282-287.
ZHANG Hong, JIANG Kai, WANG Chao, et al. The current status of SAR tomography[J]. Remote Sensing Technology and Application, 2010, 25(3): 282-287. (in Chinese)
- [5] Zhu X X, Bamler R. Very high resolution spaceborne SAR tomography in urban environment[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(12): 4296-4308.
- [6] Fornaro G, Serafino F, Soldovieri F. Three-dimensional focusing with multipass SAR data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(3): 507-517.
- [7] Li J, Stoica P. Efficient mixed-spectrum estimation with application to target feature extraction[J]. IEEE Transactions on Signal Processing, 1996, 44(2): 281-295.
- [8] Gini F, Lombardini F, Montanari M. Layover solution in multibaseline SAR interferometry[J]. IEEE Transactions on Aerospace and Electronic Systems, 2002, 38(4): 1344-1356.
- [9] 任笑真,杨汝良. 一种 SAR 层析成像的 RELAX 改进算法[J]. 数据采集与处理,2010, 25(3): 302-306.
REN Xiaozhen, YANG Ruliang. Improved RELAX algorithm for SAR tomography [J]. Journal of Data Acquisition & Processing, 2010, 25(3): 302-306. (in Chinese)
- [10] 王敏,孙希龙,余安喜,等. 基于双星编队 SAR 的差分干涉研究[J]. 国防科技大学学报,2008,30(4): 71-75.
WANG Min, SUN Xilong, YU Anxi, et al. Research on differential interferometry for bi-satellite formation SAR[J]. Journal of National University of Defense Technology, 2008, 30(4): 71-75. (in Chinese)
- [11] Sun X L, Dong Z, Yu A X, et al. Robust phase correction method for SAR tomography via phase gradient autofocus [C]//Processing of IEEE International Geoscience and Remote Sensing Symposium, 2011.
- [12] Lombardini F, Gini F. Model order selection in multibaseline interferometric radar systems [J]. EURASIP Journal on Applied Signal Processings, 2005, 20: 3206-3219.
- [13] Zhu X X, Adam N, Brcic R, et al. Space-borne high resolution SAR tomography: experiments in urban environment using TS-X data [C]// Processing of Joint Urban Remote Sensing Event, 2009: 1-8.

基于直接数据域自适应算法的相干信号 DOA 估计*

袁晓东^{1,2}, 万建伟¹, 程 翥¹, 石斌斌³, 许可¹
(1. 国防科技大学 电子科学与工程学院, 湖南 长沙 410073;
2. 中国人民解放军 94782 部队, 浙江 杭州 310021;
3. 空军预警学院 空天基预警监视装备系, 湖北 武汉 430019)

摘要: 利用直接数据域自适应算法的稳态权值代替噪声子空间构建空间谱, 构造了一种超分辨波达方向 (DOA) 估计方法。为了解决谱峰搜索时的伪峰问题, 采用参考阵元轮换得到多组线性无关稳态权值, 逼近噪声子空间, 能有效去除伪峰。针对相干信号的 DOA 估计, 进一步提出了直接数据域取对称共轭向量的解相干方法。相比子空间分解类算法, 本文算法不需估计信号源数目和协方差矩阵、不需特征分解, 复杂度仅为 $O(MP)$, 同时能有效完成解相干处理。

关键词: DOA 估计; 直接数据域; 自适应算法; 参考阵元轮换; 低复杂度

中图分类号: TN911.7 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0131-05

DOA estimation of correlated signals based on the adaptive algorithm in direct data domain

YUAN Xiaodong^{1,2}, WAN Jianwei¹, CHENG Zhu¹, SHI Binbin³, Xu Ke¹
(1. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. 94782 Force of PLA, Hangzhou 310021, China;
3. Department of Air/Space-based Earlywarning Surveillance Equipment, Air Force Radar Academy, Wuhan 430019, China)

Abstract: A novel method for high resolution direction-of-arrival (DOA) estimation is proposed. The novel approach constructs the spatial spectrum by utilizing the steady-state weights of the adaptive algorithm in direct data domain instead of the noise subspace. The proposed method can solve the false peak problem effectively in the spectral peak search. The noise subspace can be closed with several groups of linear non-correlated steady-state weights from utilizing different reference signals. A practical method for the DOA estimation of correlated signals is also presented, which utilizes the symmetric conjugate vector in direct data domain. Compared with the conventional method, the proposed approach does not need to estimate the source number and the covariance matrix. As it does not need the eigendecomposition and only has a computational complexity of $O(MP)$, it is a practical method for DOA estimation of correlated signals.

Key words: DOA estimation; direct data domain; adaptive algorithm; different reference signals; low complexity

实时高分辨波达方向 (Direction of Arrival, DOA) 估计是阵列信号处理中的重要问题。以 MUSIC 算法^[1]为代表的子空间分解类算法, 通过对阵列接收数据的数学分解 (如特征分解、奇异值分解和 QR 分解等), 将接收数据划分为信号和噪声两个相互正交的子空间, 然后利用两个子空间的正交性完成高分辨测向。但该类 DOA 估计算法, 需要协方差矩阵估计和特征分解, 通常还需要其他算法来估计空间信号源数目^[2], 计算量大。采用现有的数字信号处理芯片实现子空间分解类算法仍有难度。

现有 DOA 估计算法针对子空间分解类算法中相关问题进行了大量研究, 主要解决去除或改进协方差矩阵的估计、特征分解、信号源个数的确定以及解相干等问题, 但未能得到一种利于工程实现的解相干精确测向算法。利用直接数据域方法, 可以使用单个快拍数据进行协方差矩阵构造、噪声子空间估计等^[3-5]。文献[3]基于直接数据域最小二乘算法, 提出了一种 DOA 估计方法, 通过自适应算法得到自适应权值, 然后求取自适应权值的模之和, 搜索使模之和最小的方向即为来波方向^[3]; 文献[4]直接利用快拍数据构造 Toeplitz 矩阵代替协方差矩阵^[4], 但其仍需特征分解和确定信源数目; 文献[5]提出一种方向估计

* 收稿日期: 2011-07-21
基金项目: 国家自然科学基金资助项目 (61002022)
作者简介: 袁晓东 (1979—), 男, 河南周口人, 博士研究生, E-mail: yuanxiaodong08@nudt.edu.cn
万建伟 (通信作者), 男, 教授, 博士, 博士生导师, E-mail: kermittjw@139.com

和跟踪自适应算法,通过 LMS 算法估计噪声子空间,利用近似牛顿算法进行方向更新,但算法计算量仍很高,且不利于工程实现^[5]。

本文在分析子空间分解类算法测向原理的基础上,利用直接数据域自适应算法的稳态权值代替噪声子空间的基构建空间谱,构造了一种新的超分辨 DOA 估计方法。针对算法中谱峰搜索时出现的伪峰问题,采用参考阵元轮换得到多组线性无关稳态权值,逐步逼近噪声子空间,能有效去除伪峰。针对相干信号的 DOA 估计,提出了直接数据域取对称共轭向量的解相干算法。所提算法在减小计算量的同时,能有效完成解相干处理。

1 子空间分解类算法的测向原理

考虑 N 个远场的窄带信号从不同的方位角和俯仰角 $(\theta_1, \varphi_1), (\theta_2, \varphi_2), \dots, (\theta_N, \varphi_N)$ 入射到一个 M 元等距平面阵 ($N \leq M-1$), 阵元间距 d 为半波长 $\lambda/2$, 阵元的坐标分别为 $(x_i, y_i), (i=1, 2, \dots, M)$, 则阵列在第 k 个快拍接收的 M 维观测数据矢量为

$$x(k) = A(\theta, \varphi)s(k) + v(k) \quad (1)$$

其中:

$$A(\theta, \varphi) = [\alpha(\theta_1, \varphi_1), \alpha(\theta_2, \varphi_2), \dots, \alpha(\theta_N, \varphi_N)]^T$$

$$s(k) = [s_1(k), s_2(k), \dots, s_N(k)]^T$$

$$v(k) = [v_1(k), v_2(k), \dots, v_M(k)]^T$$

$$\alpha(\theta_j, \varphi_j) = [e_{1j}^{-j\omega\tau}, e_{2j}^{-j\omega\tau}, \dots, e_{Mj}^{-j\omega\tau}]^T$$

$$\tau_{ij} = (x_i \cos \theta_j \cos \varphi_j + y_i \sin \theta_j \cos \varphi_j) / c$$

$$i=1, 2, \dots, M; \quad j=1, 2, \dots, N$$

式(1)即为二维测向的数学模型,其中 $s(k)$ 为空间信号的 $N \times 1$ 维矢量, $v(k)$ 为 $M \times 1$ 维噪声数据矢量, $A(\theta, \varphi)$ 为空间阵列的 $M \times N$ 维导向矢量矩阵, $\alpha(\theta_j, \varphi_j)$ 为第 j 个信号的导向矢量, ω 为接收信号的中心频率, c 为信号传播速度。

假设阵列各阵元的加性噪声 $n_i(t)$ 为空间白噪声且互不相关, $i=1, 2, \dots, M$; 噪声与信号也互不相关, 各阵元为各向同性, 且不考虑互耦以及通道不一致的影响, 在这些假设条件下, 对阵列快拍数据的协方差矩阵 R_X 进行特征分解有

$$R_X = U_S \Sigma_S U_S^H + U_N \Sigma_N U_N^H \quad (2)$$

式(2)中, U_S 是由大特征值对应的特征矢量张成的子空间, 也即信号子空间, 而 U_N 是由小特征值对应的特征矢量张成的子空间, 也即噪声子空间。可以证明数据空间中的信号子空间与噪声子空间是相互正交的, 信号的导向矢量可以由信号子空间的一组基线形表示, 故信号的导向矢量

也与噪声子空间正交:

$$a^H(\theta, \varphi) U_N = 0 \quad (3)$$

子空间分解类算法就是通过搜索与噪声子空间垂直的导向矢量方向来测向的, 经典的 MUSIC 算法的谱估计公式为

$$P_{\text{MUSIC}} = \frac{1}{a^H(\theta, \varphi) U_N U_N^H a(\theta, \varphi)} \quad (4)$$

由上式得到的空间谱在信号源方向会产生较尖的“峰值”, 而在其他方向相对平坦。通过谱峰搜索即可获得空间信号的到达角。

2 自适应算法测向原理

使用第 1 节数学模型, 由上节子空间分解类算法的测向原理知, 只要得到噪声子空间, 就可以构建类似于式(4)的空间谱, 通过谱峰搜索进行测向。自适应算法中, 设阵列输入矢量为 $x(n) = [x_1, x_2, \dots, x_M]^T$, 阵列的加权矢量为 $w(n) = [w_1, w_2, \dots, w_M]^T$, 输出误差为 e 。任选一路输入做参考路, 考虑简单约束问题:

$$\min w^H R_X w \quad \text{s. t.} \quad w(i) = 1 \quad (5)$$

其中 i 对应选取的参考路。随着权值 $w(n)$ 迭代到稳态权值 w_{opt} , 阵列的最终输出能量 $E\{|e^2(n)|\} = w_{\text{opt}}^H R_X w_{\text{opt}}$ 最小, 输出中不包含信号成分, 即有: $e_{\min} = w_{\text{opt}}^H a(\theta, \varphi)$ 趋近于零, 故权值 w_{opt} 垂直于信号子空间 U_S 。由于 U_S 和 U_N 互为正交补, 故 w_{opt} 为噪声子空间内的一个矢量, 可用权值 w_{opt} 代替 U_N 构建空间谱:

$$P_{\text{ADAPT}} = \frac{1}{a^H(\theta, \varphi) w_{\text{opt}} w_{\text{opt}}^H a(\theta, \varphi)} \quad (6)$$

由于权值 w_{opt} 仅是 U_N 内的一个矢量, 故与 w_{opt} 垂直的方向矢量中既包含了信号方向, 也很可能包含了一些与 w_{opt} 垂直, 但不在 U_S 内的方向矢量, 在空间谱中表现为多余的尖峰, 本文称之为伪峰。为了去除伪峰, 需要更多不同的 w_{opt} 矢量作为基张成子空间来逼近 U_N 。

3 参考阵元轮换测向算法

选取不同阵元做参考阵元进行自适应算法时, 若阵元位置无对称关系, 则得到的 w_{opt} 线性无关, 且都属于噪声子空间 U_N 。不同的权值矢量 w_{opt} 张成 U_N 内的一个子空间, 记为 $\text{span}\{w_{\text{opt}}\}$, 则有 $\text{span}\{w_{\text{opt}}\} \subseteq U_N$ 。随着轮换参考阵元数的增多, $\text{span}\{w_{\text{opt}}\}$ 接近 U_N , 从而伪峰得到消除, 原理如图 1 所示。

图 1 中 $w_{\text{opt}1}, w_{\text{opt}2}$ 为不同参考阵元得到的自适应权值, 都位于噪声子空间 U_N 内, S_1 为所有与

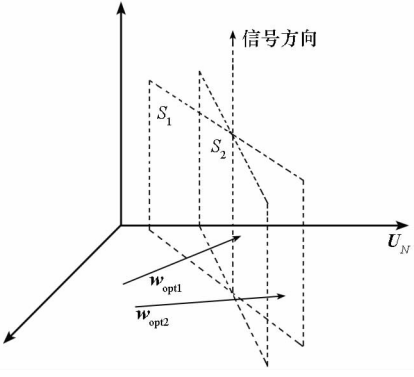


图1 参考阵元轮换自适应算法测向消除伪峰原理

Fig. 1 Principle of false peak elimination in DOA estimation

with adaptive algorithm using different reference signals

矢量 \mathbf{w}_{opt1} 垂直的方向向量张成的空间, S_2 为与矢量 \mathbf{w}_{opt2} 垂直的方向向量张成的空间。则单独搜索与矢量 \mathbf{w}_{opt1} 或 \mathbf{w}_{opt2} 垂直的方向向量不唯一, 表现在空间谱上不但包含真实信号谱峰, 也包含伪峰; 但搜索同时垂直于 $\mathbf{w}_{opt1}, \mathbf{w}_{opt2}$ 的方向向量, 则仅有 S_1 和 S_2 相交的向量, 即为信号来波方向。所以通过阵元轮换自适应算法, 最终能消除伪峰, 得到超分辨的信号方向, 参考阵元轮换自适应算法测向的空间谱为

$$P_{ADAPT} = \frac{1}{\sum_{i=1}^P a_i^H(\theta, \varphi) \mathbf{w}_{opt(i)} \mathbf{w}_{opt(i)}^H a_i(\theta, \varphi)} \quad (7)$$

其中 P 为选取的轮换参考阵元数目, 其范围为: $1 \leq P \leq M$ 。

4 基于直接数据域的解相干方法

上述算法当信号相干时会产生秩损失, 从而不能进行相干信号测向。为了解相干, 在等距线性方阵和同时相干信号个数不大于两个的情况下, 提出了直接数据域取“对称共轭向量”进行 DOA 估计。

令 \mathbf{J} 为 $M \times M$ 置换矩阵, 其除反对角线上元

$$\begin{aligned} & \begin{bmatrix} E[(s + as^*)(s + as^*)^H] & E[(s + as^*)(s + bs^*)^H] \\ E[(s + bs^*)(s + as^*)^H] & E[(s + bs^*)(s + bs^*)^H] \end{bmatrix} \\ &= \begin{bmatrix} (1 + a^2) \{E[ss^*]\} + a \{E[ss]\} + a \{E[s^*s^*]\} & (1 + ab) \{E[ss^*]\} + b \{E[ss]\} + a \{E[s^*s^*]\} \\ (1 + ab) \{E[ss^*]\} + a \{E[ss]\} + b \{E[s^*s^*]\} & (1 + b^2) \{E[ss^*]\} + b \{E[ss]\} + b \{E[s^*s^*]\} \end{bmatrix} \end{aligned}$$

其行列式经计算为: $(a^2 + b^2 - 2ab) \{E[ss^*]\}^2$, 故仅当 $a = b$, 即 $b_1 c^*/c = b_2$ 时, 行列式为 0, 此时有 $(\theta_1, \varphi_1) = (\theta_2, \varphi_2)$, c 为实数, 即两相干信号方位相同, 且无相位延迟, 可看作一个信号, 故没必要分辨; 而在其他情况下, 有 $a \neq b$, 此时矩阵行列式不为 0, 矩阵非奇异, 即两相干信号组成向量线性无关。故得到的 x_{BF} 总为满秩矩阵, 能体现出所有信源数。总结整个解相干自适应算法测向流

素为 1 外, 其余元素皆为 0, 则对于 N 个信号入射到 M 元等距线性方阵 (每行和每列阵元数都为 $L = \sqrt{M}$), 可得 \mathbf{Jx}^* 为输入快拍向量 x 的对称共轭向量。取 $x_{BF} = x + \mathbf{Jx}^*$, 新的输入快拍数据向量变为

$$\begin{aligned} x_{BF} &= \mathbf{As} + \mathbf{JA}^* s^* + \sigma^2 \mathbf{I} \\ &= \begin{bmatrix} a(\theta_1, \varphi_1) \\ \vdots \\ a(\theta_N, \varphi_N) \end{bmatrix}^T \times \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} + \begin{bmatrix} a(\theta_1, \varphi_1) \\ \vdots \\ a(\theta_N, \varphi_N) \end{bmatrix}^T \times \\ &\quad \begin{bmatrix} e^{j\pi((L-1)\sin(\theta_1)\cos(\varphi_1) + (L-1)\sin(\theta_1)\sin(\varphi_1))} s_1^* \\ \vdots \\ e^{j\pi((L-1)\sin(\theta_N)\cos(\varphi_N) + (L-1)\sin(\theta_N)\sin(\varphi_N))} s_N^* \end{bmatrix} + \sigma^2 \mathbf{I} \\ &= [a(\theta_1, \varphi_1) \cdots a(\theta_N, \varphi_N)] \times \begin{bmatrix} s_1 + b_1 s_1^* \\ \vdots \\ s_N + b_N s_N^* \end{bmatrix} \\ &\quad + \sigma^2 \mathbf{I} \\ &= \mathbf{As}_{BF} + \sigma^2 \mathbf{I} \end{aligned} \quad (8)$$

其中 $b_1 = e^{j\pi((L-1)\sin(\theta_1)\cos(\varphi_1) - (L-1)\sin(\theta_1)\sin(\varphi_1))}$, \dots , $b_N = e^{j\pi((L-1)\sin(\theta_N)\cos(\varphi_N) - (L-1)\sin(\theta_N)\sin(\varphi_N))}$ 都为复常数, N 为入射信号源数目。矩阵 \mathbf{A} 为 Vandermonde 矩阵, 当 $(\theta_1, \varphi_1), \dots, (\theta_N, \varphi_N)$ 各不相同, 其为满秩矩阵; 下面证明 s_{BF} 当同时相干信号个数不大于 2 时, 为 N 个线性无关的矢量: 易知矩阵 s_{BF} 中非相干信号组成向量线性无关。而两相干信号, 不失一般性, 设为 s_1 和 s_2 , 令 $s_1 = cs$, $s_2 = s$, 其中 c 为复常数, $s = s(t)e^{-j2\pi f_c t}$, 故两相干信号组成的向量分别为 $cs + b_1 c^* s^*, s + b_2 s^*$ 。令 $a = b_1 c^*/c$, $b = b_2$, 也都为复常数。需证明向量 $s + as^*$ 与向量 $s + bs^*$ 无关, 只需证明两者相关矩阵为满秩, 两者相关矩阵为

$$\begin{aligned} & \begin{bmatrix} (1 + a^2) \{E[ss^*]\} + a \{E[ss]\} + a \{E[s^*s^*]\} & (1 + ab) \{E[ss^*]\} + b \{E[ss]\} + a \{E[s^*s^*]\} \\ (1 + ab) \{E[ss^*]\} + a \{E[ss]\} + b \{E[s^*s^*]\} & (1 + b^2) \{E[ss^*]\} + b \{E[ss]\} + b \{E[s^*s^*]\} \end{bmatrix} \\ & \text{程为:} \end{aligned}$$

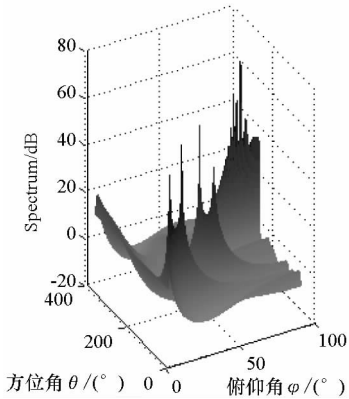
(1) 通过 $x + \mathbf{Jx}^*$ 得到新的快拍数据;

(2) 选取不同阵元做参考阵元, 经自适应算法得到不同的稳态权值 \mathbf{w}_{opt} , 一般得到两个稳态权值: $\mathbf{w}_{opt1}, \mathbf{w}_{opt2}$ 即可;

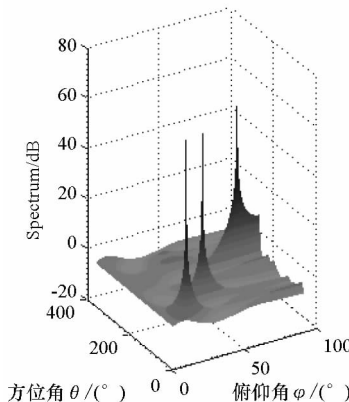
(3) 利用得到的不同稳态权值按式(7) 构造空间谱, 进行信号方向搜索。

5 仿真和测试

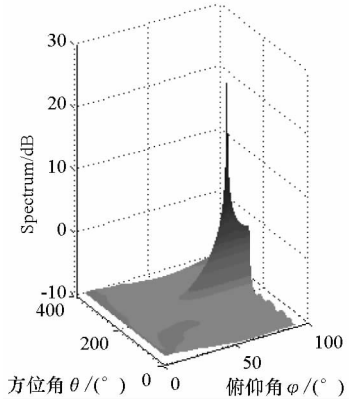
设定典型仿真条件为:阵列 4×4 均匀方阵,阵元间距为信号半波长 $\lambda/2$ 。输入信号为 3 个窄带信号,信噪比都设为 30dB。3 个信号的俯仰角和方位角分别为: $(15^\circ, 40^\circ)$, $(35^\circ, 100^\circ)$, $(75^\circ, 210^\circ)$, 其中前两个窄带信号为相干信号。分别用固定参考阵元自适应算法、参考阵元轮换自适应算法和 MUSIC 算法进行测向,空间谱结果如图 2 所示。



(a) 固定参考阵元自适应算法测向



(b) 两参考阵元轮换自适应算法测向



(c) MUSIC 算法测向

图 2 不同算法测向结果比较

Fig.2 Comparison with different DOA estimation algorithms

由图 2(a) 可看出单一稳态权值矢量 \mathbf{w}_{opt} 构建空间谱,能测出所需信号方向,但空间谱中也包含了较多的伪峰,一般可通过设置门限滤除,但在信号源数量未知时容易造成误测;图 2(b) 为用两阵元轮换做参考阵元测向结果,可看出伪峰已被有效消除,能较好地进行相干信号测向,而且算法中有取平均操作,能改善估计方差。图 2(c) 所示 MUSIC 算法仅能测出非相干信号,不能对相干信号测向。

本算法已在开发的自适应抗干扰处理机中实现,并已经投入实际应用,取得了较好的干扰定位效果。为了说明该算法的性能,本文在自适应抗干扰处理机的基础上设计了仿真实验,实验中天线阵列采用 4×4 等距方阵,AD 采样为 14 位,自适应抗干扰算法选为 LMS 算法,空间信号源设置同仿真条件,信号间非相干,实验重复 20 次,得到的信噪比对测向性能的影响以及轮换参考阵元数量 P 、信号角度间隔大小对分辨成功概率的影响分别如图 3、图 4 和图 5 所示。

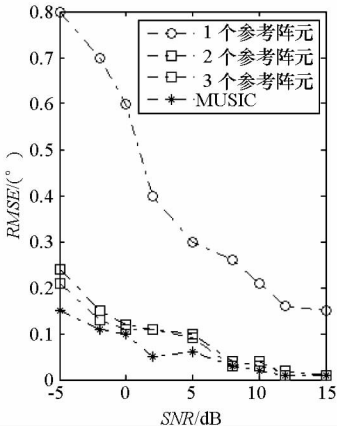


图 3 均方根误差比较
Fig.3 Comparison with RMSE

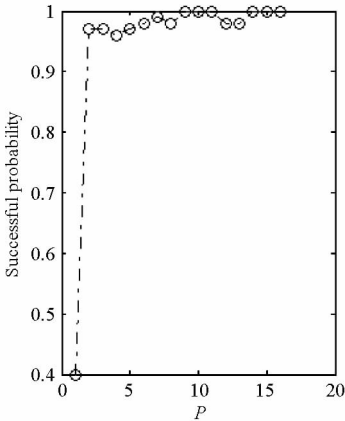


图 4 成功概率随轮换参考阵元数变化
Fig.4 Successful probability with different reference signal numbers

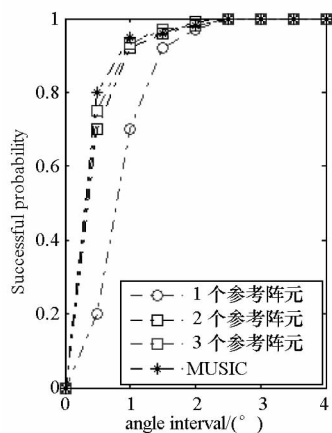


图5 角度分辨概率比较

Fig. 5 Successful probability with angle interval

由图3可看出,本文算法性能与 MUSIC 算法相当,在低信噪比条件下,由于自适应稳态权值 w_{opt} 受噪声影响较大,存在数值上的波动,导致通过本文算法测向精度比 MUSIC 算法略低;但 MUSIC 算法在低信噪比条件下存在不易划分信号子空间和噪声子空间的问题,只有信号源数目估计正确,才能得到较好测向结果。由图4可看出仅用两个不同参考阵元,即可达到较好的测向成功概率,而一个参考阵元自适应算法由于滤除伪峰的门槛较难设定,严重影响所测信号源数目结果,从而降低了分辨概率。由图5可看出随着信号角度的接近,自适应算法通过降低收敛步长仍能达到较好分辨力;MUSIC 方法通过增加接收数据采样点数,也能达到较好的分辨力。

6 结束语

本文基于自适应阵列处理算法,构造了一种用稳态权值 w_{opt} 逼近噪声子空间,从而构建空间谱进行信号 DOA 估计的新方法。深入分析了通过自适应权值测向产生伪峰,以及采用参考阵元轮换自适应算法测向可去除伪峰的原因。为解决相干信号 DOA 估计问题,采用了直接数据取“对称共轭向量”方法。经仿真和实际测试验证,其性能与直接利用噪声子空间测向类算法相当,而且其不需预知信号源数目,不需协方差矩阵估计,不需特征分解,所以极大地降低了计算量。如用

LMS 类自适应算法实现,本文算法复杂度仅有 $O(MP)$ (P 为选取的轮换参考阵元数目,一般选取 $P=2$ 即有较好结果)。因此本文所提方法能对空间相干多信号源进行精确超分辨测向,且便于工程实现。

参考文献 (References)

- [1] Schmidt R O. Multiple emitter location and signal parameter estimation [J]. IEEE Transaction. on Antennas and Propagation, 1986, 34(3): 276-280.
- [2] Wu H T, et al. Source number estimator using gerschgorin disks [C]//Proceeding of ICASSP, 1994(4): 261-264.
- [3] Azarbar A, Dadashzadeh G. A new DOA estimation based on direct data domain algorithm [C]//2011 IEEE GCC Conference and Exhibition, 2011(2): 205-208.
- [4] Chen H, Huang B, Wang Y, et al. Direction-of-arrival estimation based on direct data domain (D3) method [J]. Journal of Systems Engineering and Electronics, 2009, 20(3): 512-518.
- [5] Xin J, Sano A. Efficient subspace-based algorithm for adaptive bearing estimation and tracking [J]. IEEE Transactions on Signal Processing, 2005, 53(12): 4485-4505.
- [6] 王永良,陈辉,彭应宁,等. 空间谱估计理论与算法 [M]. 北京:清华大学出版社, 2004.
WANG Yongliang, CHEN Hui, PENG Yingning, et al. Spatial spectrum estimation theory and method [M]. Beijing: Tsinghua University Press, 2004. (in Chinese)
- [7] 程翥,李双勋,薛鸿印,等. 一种简化的单约束波束形成方法 [J]. 系统仿真学报, 2007, 19(12): 2673-2676.
CHENG Zhi, LI Shuangxun, XUE Hongyin, et al. Simplified implementation for single constrained beam-forming [J]. Journal of System Simulation, 2007, 19(12): 2673-2676. (in Chinese)
- [8] Ferreira T N, Netto S L, Diniz P S R. Direction-of-Arrival estimation using a direct-data approach [J]. IEEE Transaction. on Aerospace and Electronic Systems, 2011, 47(1): 728-733.
- [9] 于红旗,黄知涛,周一宇,等. 一种不需要特征值分解的 MUSIC 方法 [J]. 国防科技大学学报, 2007, 29(4): 91-94.
YU Hongqi, HUANG Zhitao, ZHOU Yiyu, et al. An improved MUSIC algorithm without eigenvalues decomposition [J]. Journal of National University of Defense Technology, 2007, 29(4): 91-94. (in Chinese)

基于线性频率调制与解调的片上直流电压信号放大器*

马卓, 潘国腾, 郭阳, 谢伦国
(国防科技大学 计算机学院, 湖南 长沙 410073)

摘要: 片内直流电压信号摆幅较小,且受到CMOS工艺中被动器件一致性较差、易被高频交流信号干扰等因素的影响,采用典型的比例放大电路难以放大这类信号。为此提出了一种基于“载频”的“电压→频率→电压”放大方法,使用载波信号作为片上长距离传输的信号,将易受到干扰的直流信号局部化,利用前馈补偿技术构建了具有高度线性转换关系的“电压→频率”调制电路,采用具有较高线性度的频率解调电路实现后级电压信号的解调,有效地放大片内直流电压信号。电路仿真结果表明,所提出的放大器电路能有效地放大片上电压信号,直流电压增益为2.4。

关键词: 片内直流电压信号;放大器;线性;频率调制;解调

中图分类号: TN432 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0136-06

An on-chip DC Signal Amplifier based on linear frequency modulation and demodulation

MA Zhuo, PAN Guoteng, GUO Yang, XIE Lunguo
(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: To amplify the DC signal is a discommodious issue, because of the on-chip DC signal has such properties as little swing, less coherence and likeliness to be interfered. In light of this, a “voltage-frequency-voltage” mode DC signal amplifier is proposed, in which the frequency is the carrier signal. This amplifier is highly linear, and is equipped with a high linearity “voltage-frequency” modulator and a “frequency-voltage” de-modulator. The simulation experimental results show that the DC signal amplifier amplifies the input voltage signal efficiently while achieving stable high linearity, and the DC gain is 2.4.

Key words: on-Chip DC signal amplifier; linearity; frequency modulation; demodulation

受到输出路径上的干扰及片内传输路径上的损耗,片内直流电压信号的放大输出问题一直是芯片测试中的难点。在典型的直流信号采集电路中,比例电压放大是一种十分常见的方式,图1中给出了一种典型的利用运算放大器(Operational Amplifier, OPA)进行直流电压放大的电路,利用OPA输入端“虚短(imaginary short)”的特性,闭环系统能够保证A点的电压和待放大信号 V_{in} 保持一致,输出电压与待采样信号 V_{in} 满足确定性的比例^[1],表达式给出了在基于运放的电压比例电路中的基本运算规则,其中 V_{in} 为待放大信号, V_{out} 为放大输出的电压信号, R_0 和 R_1 为比例分压电阻的阻值。

$$V_{out} = (1 + \frac{R_0}{R_1}) \cdot V_{in} \tag{1}$$

但是,Lam、Chen研究了这种结构对于输入电压信号的动态跟踪能力,指出了这种类LDO(Low

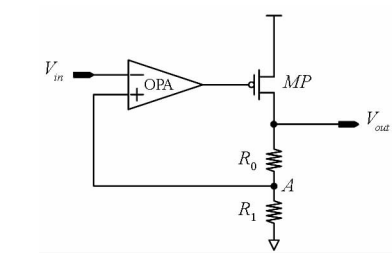


图1 基于运放的比例电压放大电路结构
Fig.1 The typical pprootional amplifier for DC voltage signal

Dropout Regulator)形式的电压放大器的跟踪能力十分有限^[9-10]。Mohaned和Xu进一步指出这种结构对电源噪声的敏感程度^[11-12]。针对这个问题,本文提出了一种“电压→频率→电压”的放大模式,并设计了具有极高线性度的“电压-频率”调制与“频率-电压”解调电路,实现片内直流电压信号的有效放大输出。信号放大过程被有效分解为调制与解调两个步骤,能有效控制信号增益;

* 收稿日期:2011-06-17
基金项目:国家自然科学基金资助项目(61076025,61070036)
作者简介:马卓(1981—),男,湖北鄂州人,助理研究员,博士研究生,E-mail:ouhzam@nudt.edu.cn

将易受干扰的直流信号有效地局部化,长距离传输使用交流信号作为“载波”,提高了放大过程的抗干扰能力;“调制”与“解调”过程均具有高线性度的转换关系,极大地提高了整体放大电路的线性化程度。

1 基于频率调制与解调的电压信号放大模型

对于直流电压信号放大而言,所实现的目的就是将待放大信号乘以一定的系数,并驱动相应的负载,这个关系如式(2)所示,其中 k 为比例系数, V_{in} 和 V_{out} 分别为待放大电压和放大后的电压信号。对应图1中的结构,式(1)中该比例系数由分压电阻 R_0 和 R_1 的阻值确定。

$$V_{out} = k \cdot V_{in} \tag{2}$$

令比例系数 $k = k_1 \cdot k_2 \cdots k_n$,则式(2)可改写为式(3)的形式。

$$V_{out} = k_1 \cdot k_2 \cdots k_n \cdot V_{in} \tag{3}$$

考虑式(3)的可实现性,定义如图2的传输系统,将比例系数 k 分解为三级比例系数,其中 k_1, k_2, k_3 为各级的比例系数。三级开环传递函数如式(4)或(5)所示。

$$u_o = k_3 \cdot u_e = k_3 \cdot k_2 \cdot u_c = k_3 \cdot k_2 \cdot k_1 \cdot u_i \tag{4}$$

$$u_o = k_3 \cdot t_e = k_3 \cdot k_2 \cdot u_c = k_3 \cdot k_2 \cdot k_1 \cdot u_i \tag{5}$$

第一种方式为信号传播方式不变的放大过程,即图2中第二级 k_2 的输出保持 u_e 的形式和物理信号载体,在这种方式下,三级开环传输系统中信号的物理形式保持不变。另一种方式是在中间级的传输过程中使用其他信号物理形式,如图2中第二级 k_2 的输出使用了 t_e 的形式,这正是一种应用“调制-解调”技术的传输模型。

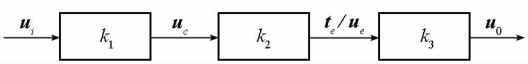


图2 用于比例放大的三级开环传输系统
Fig.2 Triple stages transform system for propotional amplify

对于第一种方式而言,由于信号的物理形式在传输过程中没有发生改变,因此该模型适合高电压增益的比例信号放大。但是在复杂系统芯片(System-on-Chip, SoC)中,供电电压十分有限,并且信号传递路径上可能存在复杂的电磁环境,信号传递路径较长,在信号放大的中间级使用适合的信号物理形式显得十分重要,第二种方式更为有效。

2 压控频率调制

应用于电压信号和频率信号之间的信号调制问题是一个典型的电压控制产生振荡的问题,可

以采用压控振荡器来实现。图3(a)是典型的电压控制振荡模型,图3(b)是一种常见的压控环形振荡器(Voltage Controlled Oscillator, VCO)结构,能够将控制电压调制成为交流振荡信号^[2]。相对于差分结构的VCO而言,图3这种单端结构的VCO结构更为简单,并且具有更宽范围的输出频率^[3-4],适合载波调制领域的应用。

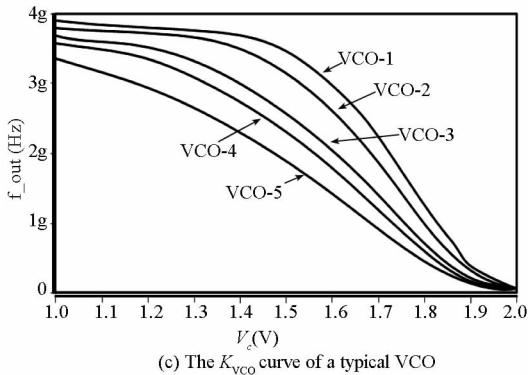
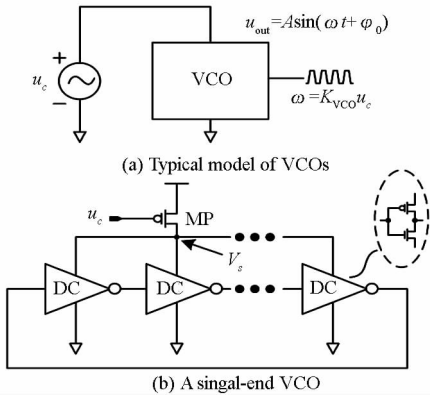


图3 典型的电压控制环形振荡器模型及其 K_{VCO} 曲线

Fig.3 A typical model, structure and K_{VCO} curve of voltage controlled oscillator

但是,图3中的典型VCO结构却不适合电压信号测量领域,式(16)给出了这种环形振荡器中单元延迟与控制电压的关系^[5],从式(16)中可见环路的振荡频率 ω 与环形振荡器的输入电压 V_s 之间存在非线性的对应关系,其中 C_{load} 为延迟单元的负载电容, λ 为晶体管的沟长调制系数, t_d 为单级延迟单元的延迟。同时,用于提供 V_s 的调整管MP也是一种非线性元件,因此环路的振荡频率 ω 与VCO的控制电压 u_c 表现为复杂的高阶非线性特性,这种高阶的非线性如图3(c)中频率与电压的关系曲线,其中VCO-1~VCO-5分别为图3(b)中的VCO采用不同电路参数的实现结果。

$$\frac{1}{\omega} = n \cdot t_d = n \cdot (-0.4 \cdot C_{load} \cdot \lambda \cdot V_s^2 + 0.5 \cdot \frac{C_{load}}{I_D} \cdot V_s) \tag{6}$$

对于信号测量而言,非线性的负面影响是显而易见的。VCO 环振中连接两级反相延迟单元的电路节点的电压变化可以等效为 RC 电路的充放电过程,因此用图解的方法能够非常清晰地展现出这种非线性的形成过程。图 4 中给出了这种非线性的关系,其中曲线 $V_{s1} \sim V_{s10}$ 表示在不同 V_s 时前述节点电压随时间的变化曲线。对图 4 (a) 中所有节点电压达到对应 V_s 的 90% 的点拟合曲线 Cur,即为电压与周期的曲线。进一步地,可以推演出图 4(b) 中频率 ($1/t$) 与电压 V 的关系曲线,显然这条曲线具有复杂的非线性关系。换

言之,如果用上述这种环形振荡器,仅可能在一个很小的控制电压范围内实现“近似的”线性。

事实上,图 4(b) 中的这种复杂的非线性关系可以通过一种前馈修正技术来解决^[6-7]。以图 5 中的延迟链为例, $D_0 \sim D_9$ 组成了一个由 10 级反相延迟单元构成的延迟链,同时传输门结构的延迟单元 T^* 和 T^{**} 分别对节点 C 施加来自节点 A 和节点 B 的前馈信号。节点 C 的电平翻转过程受到节点 A 和节点 B 的修正作用,使得时间常数 τ 与充电电流的关系趋于线性化。

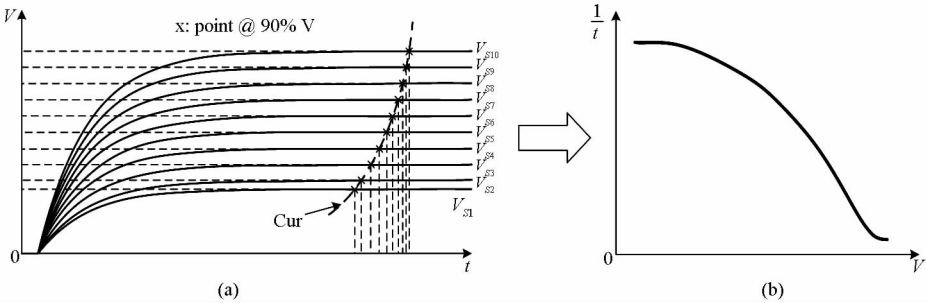


图 4 典型环振的“频率-电压”响应关系
Fig. 4 The relation curve between the frequency and voltage of the typical VCO

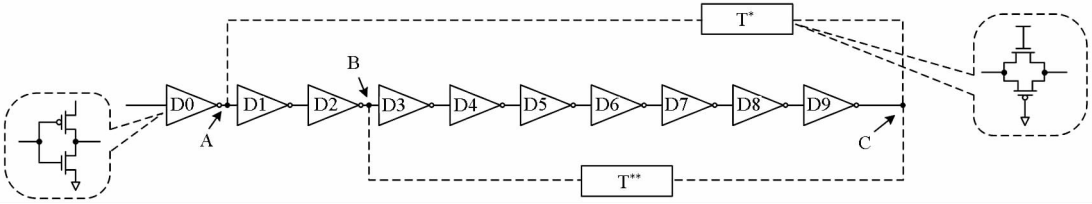


图 5 单环前馈结构中的信号通路
Fig. 5 The signal path in A single ring

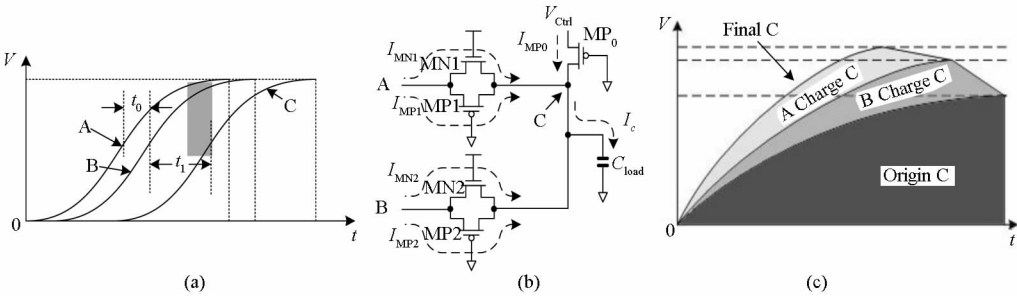


图 6 节点 C 的 $V-t$ 关系的线性化过程
Fig. 6 Linearity on the $V-t$ curve at the node C

进一步地,图 6 展示了上述的线性化过程。图 6(b) 中 C_{load} 为前述节点的等效负载电容, I_c 为该过程中对 C_{load} 充电的电流。很显然,图 5 中节点 A、B、C 的相位具有一定的顺序关系。对于图 5 中的主延迟链 $D_1, D_2 \cdots D_9$ 而言,在没有前馈 T^* 和 T^{**} 通路的情况下, A、B、C 节点均只受本级延迟单元的控制,且相位关系如图 6(a) 所示,其中节点 A 与节点 B 的相位差为 t_0 , 节点 B 与节点 C

的相位差为 t_1 , 各个节点均按照相似的充电/放电过程完成信号翻转。当考虑 T^* 和 T^{**} 前馈通路后,如图 6(b) 所示,节点 A 和节点 B 分别通过与各自相连的传输门向节点 C 补充电流,使得节点 C 的电压处于一种超指数关系上升的状态,图 6 (c) 给出了这个过程。因此,可以作出在具有前馈电流补充的情况下的节点电压曲线如图 7(a) 所示。与图 4(a) 中的 Cur 曲线对比, 图 7(a) 中

的 Cur 曲线具有更接近于双曲率曲线的关系,因而“频率-电压”具有趋近于线性的对应关系。这种对应关系与线性关系的逼近程度受到具体的

电路参数的影响。对于节点 C 的放电阶段,也可以有类似的线性化过程。

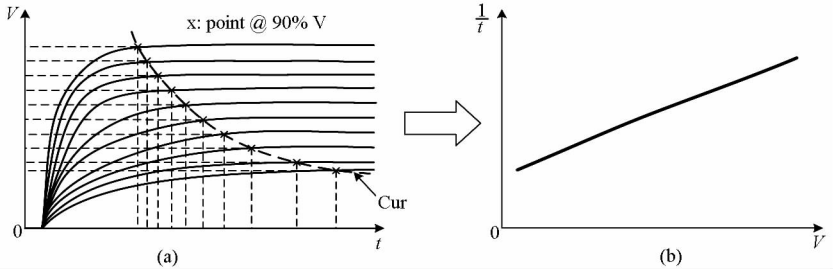


图 7 节点电压的变化“频率-电压”响应关系
Fig. 7 The relation curve between the frequency and voltage with feedforwarding

由于“电压→频率”调制过程具有接近线性的转换关系,因此该类型的压控振荡器可以完成图 2 中 k_2 增益级的调制与信号放大功能,其增益表现为 VCO 的增益 K_{VCO} 。

3 电压信号解调

相对于“电压→频率”的调制过程,将信号还原的“频率→电压”解调也需要满足线性转换的要求,以满足信号测量的易用性要求。常见的,基于电容的电流积分方法可以实现频率到电压的转换。Hung 给出了一种利用电容电荷充分分配原理的“频率→电压”转换电路,其转换电路模型如图 8 所示^[8]。不妨假设开关 S_0 和 S_1 工作互斥,且输入时钟的正半周开关 S_0 闭合,负半周开关 S_1 闭合,则当第一次时钟正半周出现的时候,电容 C_0 和 C_1 均被充电,其后随着时钟负半周到来,电容 C_0 的电量会被下拉电流源 I_D 释放一部分(释放的电量与 S_1 的闭合时间相关),当随后的时钟正半周出现时,电流源 I_C 给电容充电的同时,电容 C_0 和 C_1 还会通过闭合的开关 S_0 进行电荷充分分配,经过多次迭代之后,电容 C_1 非接地端的电位将保持稳定。

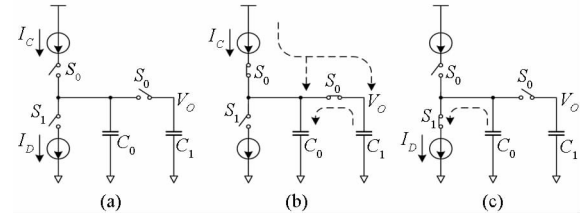


图 8 基于电荷充分分配的“频率→电压”转换模型
Fig. 8 A “Frequency → Voltage” transformative model
整体结构。该电路由前置放大模块 (PreAmplify, PA)、“电压→频率”调制模块 (Interleaved Poly-Loop VCO, IPL-VCO) 和“频率→电压”解调模块 (Frequency-Voltage Demodulator, F2V) 组成。根据第 2 节给出的技术思路,这 3 个模块分别完成 k_1 、 k_2 、 k_3 3 个阶段的信号放大。

基于第 3 节的方法,为了满足电压信号测量应用中对于 VCO 线性度的需要,同时降低对被检测信号预处理的要求,可以构建一种利用环间前馈耦合技术的多环自交叉 VCO (IPL-VCO)。图 9 (b) 部分是 IPL-VCO 的具体结构形式,16 级反相延迟单元 (D_0 、 D_1 、 $D_2 \cdots D_{15}$) 首尾相连构成环形结构,而其中的任意 9 级延迟链通过常导通传输门结构的延迟单元 (T_0 、 T_1 、 $T_2 \cdots T_{15}$) 构成子环路,即在整个结构中存在 1 个由 16 级反相延迟单元构成的主环形链和 16 个由 9 级反相延迟单元和 1 级传输门结构的延迟单元构成的子环形链。

根据 Hung 的分析,如果令充电电流 I_C 与放电电流 I_D 相等,在经过足够长的充放电迭代之后,图 8 中电路的输出电压 V_o 将满足式(7)给出的关系,其中 N 为迭代次数, t 为时钟周期。随着迭代次数的不断增加,输出电压将固定为式(8)给出的结果,很显然这也是一种线性的转换关系。

在 IPL-VCO 中,任何一个连接两个反相延迟单元的电路节点除了受到本级反相延迟单元 D^* 的作用外,还受到两路来自传输门结构延迟单元的作用,也就构成了与图 5 中延迟链相同的结构。根据第 3 节的分析,IPL-VCO 的“频率-电压”线性度很高。图 10 (a) 是 65nm 工艺下实现的 IPL-VCO 的 K_{VCO} 曲线,在 0.8 ~ 2.0 V 的电压范围内,IPL-VCO 的“频率-电压”响应具有很高的线性度。

$$V_o = \frac{I_C \cdot t}{C_1} \cdot \sum_{i=1}^N \left(\frac{C_1}{C_1 + C_0} \right)^i \tag{7}$$
$$V_o = \frac{I_C \cdot t}{C_1} \tag{8}$$

4 电路与仿真结果

图 9 是本文提出的直流电压信号放大电路

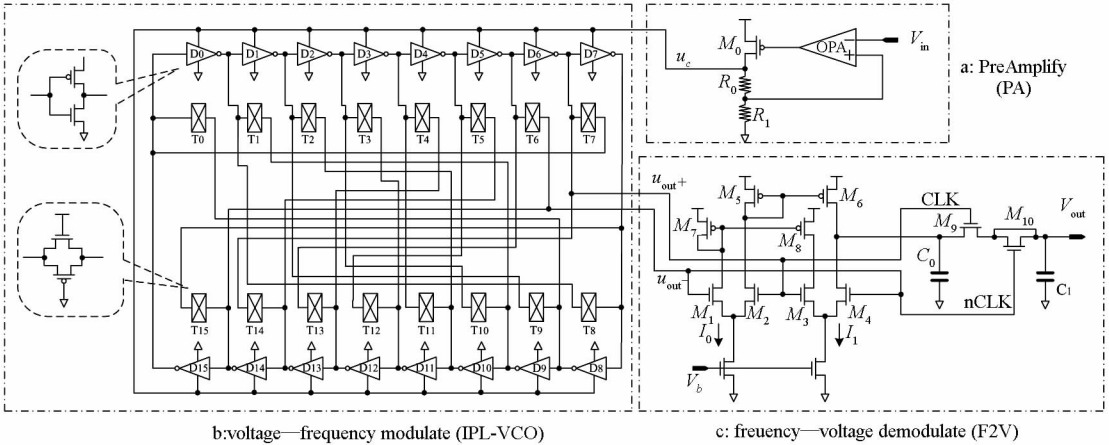


图 9 直流电压信号测试电路整体结构
Fig.9 The circuit of the whole DC signal amplifier

所示。

图 10 中整体电路的输入输出电压情况如图 11 中曲线所示,其中纵坐标为电压值,横坐标为测试样本,“□”为输入电压 V_{in} ,“Δ”为输出电压 V_{out} 。对输出电压 V_{out} 的拟合曲线表明,本文给出的直流电压信号放大器能有效地实现直流电压信号的线性放大,放大倍数可达 2.4。

从逻辑功能的角度分析,图 9 中的 3 个模块共同实现了直流电压信号放大,逻辑上 3 个模块具有紧密的承接关系。但从芯片内部电路的拓扑结构来看,由于载波信号不易被干扰,因此用于“频率(电压)解调的 F2V 模块可以远离被检测信号,而只是将 PA、IPL-VCO 模块与被检测信号局部化,甚至将 F2V 模块置于片外,利用片外的高电源电压实现更大的放大倍数,并有效解决放大过程中的噪声问题。

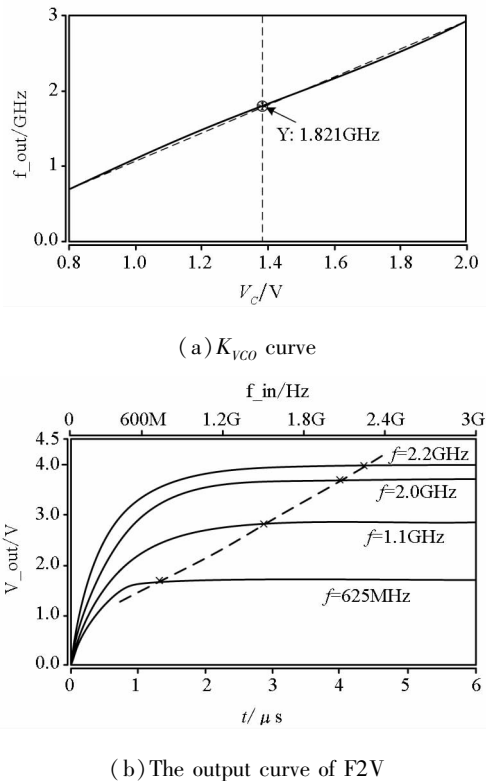


图 10 IPL-VCO 的 K_{VCO} 曲线和“频率→电压”转换电路的输出曲线

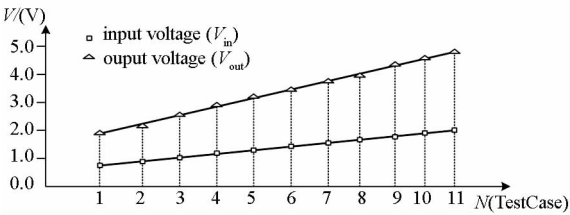


图 11 各种直流输入下放大电路的输出
Fig. 11 The Input/Output relation of the whole circuit

5 结 论

图 9(c) 部分是根据第 4 节 Hung 提出的方法实现的“频率→电压”解调电路 F2V。MOS 管 M5 ~ M8 构成的对称电流镜结构保证了充放电电流的均衡。图 10(b) 是该电路在 3 个典型输出频率下的输出电压曲线,在经历足够长的时间之后,输出电压可以达到稳定值,且稳定输出电压与输入频率之间存在较为线性的对应关系。

此外,一般情况下,由于被检测的电压信号缺乏足够的驱动能力,必须使用预放大模块 PA 进行电压跟踪,完成前级放大,并供给 IPL-VCO 所需的工作电流,PA 的电路结构如图 9(a) 部分

直流信号的放大问题是集成电路测试中的一个难点,传统的比例放大方法不能有效消除信号传播路径上的干扰。本文提出了一种具有极高线性度的直流电压信号放大电路,基于“电压→频率→电压”的分阶段放大技术,利用频率作为信号的载体,实现对直流电压信号的有效放大,并且由于“电压→频率”调制过程和“频率→电压”解

调过程均具有极高的线性度,被放大的信号具有良好的可观测性。

参考文献 (References)

[1] Shao Y L, Wang Y, Ning Z H, et al. Analysis and design of high power supply rejection LDO[C]// 8th IEEE International Conference on ASIC, ASICON 2009, October 20 – 23, 2009, Changsha, China; IEEE Computer Society, 2009; 324 – 327.

[2] Wismar U, Wisland D, Andreani P. Linearity of bulk-controlled inverter ring VCO in weak and strong inversion [C]// NORCHIP Conference, 2005; 145 – 148.

[3] 赵振宇,郭斌,张民选,等. 一款 0.18 μm CMOS 辐射加固差分压控振荡器[J]. 国防科技大学学报, 2009, 31(6): 12 – 17.

ZHAO Zhenyu, GUO Bin, ZHANG Minxuan, et al. A radiation-hardened-by-design differential voltage-controlled oscillator implemented in 0.18 μm CMOS process [J]. Journal of National University of Defense Technology, 2009, 31(6): 12 – 17. (in Chinese)

[4] 赵振宇,蒋仁杰,张民选,等. 差分压控振荡器中单粒子瞬变的研究[J]. 国防科技大学学报, 2009, 31(2): 81 – 85.

ZHAO Zhenyu, JIANG Renjie, ZHANG Minxuan, et al. Research on single-event transients in differential voltage-controlled oscillators [J]. Journal of National University of Defense Technology, 2009, 31(2):81 – 85. (in Chinese)

[5] Jan M R, Borivoje N. Digital integrated circuits—A design perspective[M]. Prentice Hall, 2003.

[6] Sun L Z, Kwasniewsk T A, A 1.25GHz 0.35 μm monolithic CMOS PLL based on a multiphase ring oscillator[J]. Solid-

State Circuits, IEEE Journal, 2001, 36(6): 910 – 916.

[7] Sun L Z, Kwasniewsk T A. A quadrature output voltage controlled ring oscillator based on three-stage sub-feedback loops [C]// Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on, 1999,172; 176 – 179.

[8] Hung T B, Yvon S. Design of a high-speed differential frequency-to-voltage converter and its application in a 5GHz frequency-locked loop[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2008, 55(3): 766 – 774.

[9] Lam Y H, Ki W H. A 0.9v 0.35m adaptively biased CMOS IDO regulator with fast transient response [J]// Proceedings of 2008 IEEE International Solid State Circuits Conference, ISSCC, Feb 3 – 7 2008, San Francisco, CA, United States, 2008: 442 – 443.

[10] Chen H, Leung K N, A fast-transient LDO based on buffered flipped voltage follower [C]// Electron Devices and Solid-State Circuits (EDSSC), 2010 IEEE International Conference, 2010; 1 – 4.

[11] Mohaned E N, Ahmed A. A 25mA 0.13 μm CMOS LDO regulator with power-supply rejection better than –56dB up to 10MHz using a feedforward ripple-cancellation technique [C]// IEEE International Solid-State Circuits Conference. ISSCC 2009. 2009; 330 – 331,331a.

[12] Xu G, Jiang J G, Wang J K. A wide range high power supply rejection ratio and transient enhanced low drop-out regulator [C]// 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), 2010; 1 – 4.

协同 OFDM 放大前传空时编码系统中的功率分配*

王丽洁¹, 唐 艺¹, 黄勤飞², 魏急波¹

(1. 国防科技大学 电子科学与工程学院, 湖南 长沙 410073;
2. 中国电子系统设备工程总公司, 北京 100141)

摘要: 针对由源节点和单个中继节点形成的协同正交频分复用 (OFDM) Alamouti 空时块码系统, 分析了放大前传 (AF) 模式下的系统容量, 给出了解析表达式。考虑节点功率独立约束的应用场景, 提出了分步迭代的功率分配算法 (SIPAA) 来提高系统容量。该算法的核心思想是在每次迭代中, 分步对源和中继节点分别进行注水功率分配, 注水功率分配前对目标函数进行线性化处理, 功率分配后对功率分配向量进行线性化处理以降低算法复杂度。仿真结果表明, 相对于等功率分配算法, 所提出的分步迭代功率分配算法可以显著提高系统容量, 且仅需很少几次迭代运算就可以完成功率分配。

关键词: 正交频分复用; 空时块码; 协同传输; 放大前传; 注水功率分配;
中图分类号: TN929.5 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0142-06

The power allocation of cooperative OFDM-based amplify-and-forward STBC system

WANG Lijie¹, TANG Yi¹, HUANG Qinfei², WEI Jibo¹

(1. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. China Electronic System Engineering Company, Beijing 100141, China)

Abstract: A cooperative orthogonal frequency division multiplexing (OFDM) system with a source and a non-regenerative relay transmitting according to Alamouti's STBC is considered. The capacity of the system is analyzed and the analytical expression is presented. To improve the capacity of the individual power constrained system, a power allocation algorithm named as the Stepwise Iterative Power Allocation Algorithm (SIPAA) is proposed. The basic principle of SIPAA is that the water-filling power allocation is implemented stepwise for both the source and the relay within iterations. To reduce the complexity of SIPAA, the object function is linearly processed before the water-filling power allocation, and the power vector is also linearly processed after the water-filling power allocation. It is confirmed by the simulation results that a significant system capacity improvement can be achieved by the proposed algorithm, compared with conventional equal power allocation algorithm. Moreover, the power allocation can be completed within a few iterations.

Key words: orthogonal frequency division multiplexing; space-time block coding; cooperative transmission; amplify-and-forward; water-filling power allocation

多天线传输的空时块码 (STBC)^[1-2] 技术在不牺牲额外带宽下, 能够显著改善系统性能。由于体积限制, 小型终端通常无法架设多根天线, 通过多个单天线以协同的方式构成一个分布式的多天线系统, 因此可以有效地利用 STBC 技术, 获得多阶的分集增益^[3-4]。协作过程中, 中继节点首先接收源节点发送的信息符号, 然后将接收到的信息进行 AF 或解码前传 (DF)。AF 下中继节点仅需将接收到的信号进行简单的放大, 其实现简单, 对设备要求低, 因此在协同通信系统中得到了广泛的关注。

分布式 STBC 协同系统对同步的要求很高, 而同步技术是无线通信中的一个难点, 参与传输的中继节点越多, 同步越难实现。文献[5-6]中给出由源节点和一个中继节点构成的采用 Alamouti 空时块码的协同模型, 由于只有一个中继协作, 其更具实现性。该模型中, 协同传输分两个阶段。第一阶段, 源节点发送信息给中继节点; 在第二阶段, 中继节点和源节点依据 Alamouti 空时码的码字结构同时向目的节点发送信息。

宽带通信中, 信道将呈现出频率选择性衰落, OFDM 技术能够有效地对抗频率选择性衰落。在

* 收稿日期: 2011-07-01
基金项目: 国家自然科学基金资助项目 (61002032, 60903206); 教育部高等学校博士点专项科研基金资助项目 (2009430711004)
作者简介: 王丽洁 (1981—), 女, 陕西渭南人, 博士研究生, E-mail: lilywang_04601@163.com;
魏急波 (通信作者), 男, 教授, 博士, 博士生导师, E-mail: wjbhw@nudt.edu.cn

频率选择性信道中,每个子载波经历不同的衰落,将总功率在每个子载波上按其信道增益进行注水功率分配可最大化系统容量^[7]。然而,在基于 OFDM 的协同通信系统中,信息传输分两个阶段进行,其功率分配与单跳传输的功率分配不完全相同。高信噪比(SNR)下,文献[8]首先获得 AF 和 DF 下的每个子载波对端到端的等价信道增益,后在总功率的约束下在每个子载波上进行注水的功率分配。对于节点功率独立受限的采用 AF 的协同系统,文献[9-10]交替的对源节点和中继节点分别进行注水的功率分配来提高系统容量。

就作者所知,目前对协同 OFDM STBC 系统的功率分配研究还很少。本文研究节点功率独立受限时,协同 OFDM STBC 系统^[6]中基于即时信道增益的功率分配。首先分析给出系统容量的解析表达式,然后在此基础上提出了分步迭代算法(SIPAA)来逐步提升系统容量。SIPAA 的每次迭代中,依次固定另两组功率分配向量,分别对源节点第1阶段、第2阶段的功率分配向量,中继节点第2阶段功率分配向量进行注水分配。为降低算法的复杂度,功率分配前,对目标函数进行线性化处理来满足接收端信噪比是功率分配向量与等价信道增益的乘积;注水功率分配后,对每个子载波上分配的功率进行线性比例化以满足总功率约束。

1 系统模型

考虑三节点协同传输模型,源节点 S 在中继节点 R 的协助下向目的节点 D 发送信息。假设每个节点安装单根天线,工作在半双工模式,即不能同时发送和接收。物理层采用 OFDM 技术,数据在 N 个正交子载波上传输。为简化分析,假设每个子载波都为单位带宽,所有节点之间的信道都是相互独立的瑞利信道,信道状态在一帧内保持不变。一个完整的信息传输分为两个阶段:第1阶段,源节点广播信息;第2阶段,源节点和中继节点采用 Alamouti STBC 同时向目的节点发送信息。目的节点只接收第2阶段的信息。

在第1个阶段,两个连续的 OFDM 符号周期里,源节点在第 k 个子载波上发送单位能量符号 $x_i, i=1,2$ 。中继节点在第 k 个子载波上接收到的信号 r_i 为

$$r_i = \sqrt{P_k^{S1}} h_k^{SR} x_i + n_{k,i}^R, \quad i=1,2 \quad (1)$$

其中 P_k^{S1} 为第1阶段 S 在第 k 个子载波上的发送

功率。 h_k^{SR} 为 $S-R$ 链路第 k 个子载波上的信道参数,其对应的信道增益为 $H_k^{SR} = |h_k^{SR}|^2$ 。Rayleigh 信道下, H_k^{SR} 服从均值 $d_{SR}^{-\alpha}$ 的负指数分布, d_{SR} 为 $S-R$ 间的距离, α 为路径损耗系数。 $n_{k,i}^R$ 为零均值且方差为 N_0 的加性高斯白噪声。

在第2个阶段,源节点和中继节点使用 Alamouti STBC 进行传输,对应的码字为

$$\begin{bmatrix} x_1 & -r_2^* \\ x_2 & r_1^* \end{bmatrix} \quad (2)$$

中继节点对接收到的信息(码字第2列)进行放大前传,源节点则重复发送第1阶段的信息(码字第1列)。因此,目的节点在第 k 个子载波上接收到的信号为

$$y_1 = \sqrt{P_k^{S2}} h_k^{SD} x_1 + \beta_k \sqrt{P_k^{S1}} h_k^{SR*} h_k^{RD} x_2^* - \beta_k h_k^{RD} n_{k,2}^R + n_{k,1}^D \quad (3)$$

$$y_2 = \sqrt{P_k^{S2}} h_k^{SD} x_2 + \beta_k \sqrt{P_k^{S1}} h_k^{SR*} h_k^{RD} x_1^* + \beta_k h_k^{RD} n_{k,1}^R + n_{k,2}^D \quad (4)$$

其中 $\beta_k = \sqrt{P_k^R / (P_k^{S1} H_k^{SR} + N_0)}$ 为 R 在第 k 个子载波上的放大系数。 P_k^R 和 P_k^{S2} 分别为第2阶段 R 和 S 在第 k 个子载波上的发送功率。 h_k^{SD} 和 h_k^{RD} 分别为 $S-D$ 链路和 $R-D$ 链路在第 k 个子载波上的信道参数,其对应的信道增益分别为 H_k^{SD} 和 H_k^{RD} 。Rayleigh 信道下 H_k^{SD} 和 H_k^{RD} 分别服从均值为 $d_{SD}^{-\alpha}$ 和 $d_{RD}^{-\alpha}$ 的负指数分布,其中 d_{SD} 和 d_{RD} 分别为 $S-D$ 和 $R-D$ 之间的距离。 $n_{k,i}^D, i=1,2$ 为零均值且方差为 N_0 的加性高斯白噪声。

假设目的节点已知各传输阶段的信道状态信息,则接收到的 x_1 和 x_2 判决表达式分别为

$$\tilde{x}_1 = \sqrt{P_k^{S2}} h_k^{SD*} y_1 + \beta_k \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} y_2^* \quad (5)$$

$$\tilde{x}_2 = \sqrt{P_k^{S2}} h_k^{SD*} y_2 - \beta_k \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} y_1^* \quad (6)$$

将(3)式和(4)式代入(5)式和(6)式,判决表达式可以重新表示为

$$\begin{aligned} \tilde{x}_1 = & x_1 (P_k^{S2} H_k^{SD} + \beta_k^2 P_k^{S1} H_k^{RD} H_k^{SR}) \\ & - \beta_k \sqrt{P_k^{S2}} h_k^{SD*} h_k^{RD} n_{k,2}^{R*} \\ & + \sqrt{P_k^{S2}} h_k^{SD*} n_{k,1}^D + \beta_k^2 \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} n_{k,1}^R \\ & + \beta_k \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} n_{k,2}^{D*} \end{aligned} \quad (7)$$

$$\begin{aligned} \tilde{x}_2 = & x_2 (P_k^{S2} H_k^{SD} + \beta_k^2 P_k^{S1} H_k^{RD} H_k^{SR}) \\ & + \beta_k \sqrt{P_k^{S2}} h_k^{SD*} h_k^{RD} n_{k,1}^{R*} \\ & + \sqrt{P_k^{S2}} h_k^{SD*} n_{k,2}^D + \beta_k^2 \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} n_{k,2}^R \\ & - \beta_k \sqrt{P_k^{S1}} h_k^{RD} h_k^{SR*} n_{k,1}^{D*} \end{aligned} \quad (8)$$

令 $A = P_k^{S2} H_k^{SD} + \beta_k^2 P_k^{S1} H_k^{RD} H_k^{SR}$, 对判决后的符号进行最大似然检测,其检测的表达式为

$$\hat{x}_1 = \arg \min_{\hat{x}_1 \in C} (|\tilde{x}_1 - A\hat{x}_1|) \quad (9)$$

$$\hat{x}_2 = \arg \min_{\hat{x}_2 \in C} (|\tilde{x}_2 - A\hat{x}_2|) \quad (10)$$

其中 C 是调制符号的集合。

2 分布迭代的功率分配

基于(7)、(8)式可得判决后两符号在第 k ($1 \leq k \leq N$) 个子载波上的接收 SNR 相同,为

$$\gamma_k = \frac{(P_k^{S2} H_k^{SD} (P_k^{S1} H_k^{SR} + N_0) + P_k^R H_k^{RD} P_k^{S1} H_k^{SR})}{((P_k^{S1} H_k^{SR} + N_0) + P_k^R H_k^{RD}) N_0} \quad (11)$$

系统容量为 N 个子载波的信道容量之和,表示为

$$C = \sum_{k=1}^N \frac{1}{2} \log_2(1 + \gamma_k) \quad (12)$$

其中因子 $1/2$ 是由于协同传输是由两个阶段组成的。

对节点功率独立受限的系统,将(11)式代入(12)式,最大化系统容量就等价于

$$C^d = \arg \max_{P_{S1}, P_{S2}, P_R} \frac{1}{2} \sum_{k=1}^N \log_2 \left(1 + \left(\frac{P_k^{S2} H_k^{SD} P_k^{S1} H_k^{SR} + P_k^R H_k^{RD} P_k^{S1} H_k^{SR} + P_k^{S2} H_k^{SD} N_0}{(P_k^{S1} H_k^{SR} + P_k^R H_k^{RD} + N_0) N_0} \right) \right) \quad (13)$$

$$\text{s. t.} \quad \mathbf{1}^T \mathbf{P}_{S1} \leq P^{S1} \text{ 且 } \mathbf{P}_{S1} \geq 0$$

$$\mathbf{1}^T \mathbf{P}_{S2} \leq P^{S2} \text{ 且 } \mathbf{P}_{S2} \geq 0$$

$$\mathbf{1}^T \mathbf{P}_R \leq P^R \text{ 且 } \mathbf{P}_R \geq 0$$

其中 P^{S1} 、 P^{S2} 分别为源节点在第 1 阶段、第 2 阶段的最大总发送功率, P^R 为中继节点的最大发送功率。定义 $\mathbf{P}_{S1} = [P_1^{S1}, P_2^{S1}, \dots, P_N^{S1}]^T$, $\mathbf{P}_{S2} = [P_1^{S2}, P_2^{S2}, \dots, P_N^{S2}]^T$, $\mathbf{P}_R = [P_1^R, P_2^R, \dots, P_N^R]^T$ 分别为每个子载波上的发送功率向量。

与单跳的 OFDM 系统不同,协同 STBC 系统容量不但依赖于 $S-D$ 链路在每个子载波上的信道增益,也依赖于 $S-R$ 和 $R-D$ 链路在每个子载波上的信道增益。从(13)式可看出源节点第 1 阶段、第 2 阶段的功率分配及中继节点的功率分配互相影响并制约着系统容量。在任意两条链路功率分配给定的情况下,总能通过注水算法得到第 3 条链路的最优功率分配。

基于上述分析,本文提出一个低复杂度的分步迭代注水算法。算法执行之前,首先平均功率分配初始化源节点和中继节点每个信息传输阶段在每个子载波上的功率,即对任意第 k 个子载波, $P_{k(0)}^{S1} = P^{S1}/N$, $P_{k(0)}^{S2} = P^{S2}/N$, $P_{k(0)}^R = P^R/N$ 。通过对(11)式分析可得, γ_k 是 P_k^{S1} , P_k^{S2} , P_k^R 的增函数,

因此系统容量最大化的情况下,源节点和中继节点的总功率将被全部使用。第 j 次迭代由以下 3 步构成。

第 1 步: $S-R$ 链路的注水功率分配

固定向量 $\mathbf{P}_{S2}^{(j-1)}$ 和 $\mathbf{P}_R^{(j-1)}$, 最大化系统容量的功率分配问题表示为

$$C_{(j)}^d = \arg \max_{\mathbf{P}_{S1}^{(j)}} \frac{1}{2} \sum_{k=1}^N \log_2 \left(1 + \left(\frac{P_{k(j-1)}^{S2} H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^{S2} H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0} \right) \right) \quad (14)$$

$$\text{s. t.} \quad \mathbf{1}^T \mathbf{P}_{S1}^{(j)} = P^{S1} \text{ 且 } \mathbf{P}_{S1}^{(j)} \geq 0$$

(14) 式中,每个子载波的 SNR 项中 $P_{k(j)}^{S1}$ 同时存在于分子和分母中,这使注水功率分配的计算复杂化。为简化计算,分母中的 $P_{k(j)}^{S1}$ 用 $P_{k(j-1)}^{S1}$ 代替,在计算分子的 $P_{k(j)}^{S1}$ 时, $P_{k(j-1)}^{S1}$ 固定不变。在(14)式的目标函数中,如果第 $j-1$ 次迭代后第 k 个子载波 $P_{k(j-1)}^{S2} = 0$ 且 $P_{k(j-1)}^R = 0$, 则第 j 次迭代中源节点在第 1 个传输阶段给这个子载波分配功率失去意义。寻找所有功率分配无意义的子载波,设这些子载波上的功率分配值为 0,然后在剩余的子载波上注水功率分配。

$$\text{令 } a_{k(j)} = \frac{P_{k(j-1)}^{S2} H_k^{SD} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} H_k^{SR}}{(P_{k(j-1)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0},$$

$$b_{k(j)} = \frac{P_{k(j-1)}^{S2} H_k^{SD}}{P_{k(j-1)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0}, \quad (14) \text{ 式最优}$$

$$\text{化问题的目标函数就表示为}$$

$$C_{(j)}^d = \arg \max_{\mathbf{P}_{S1}^{(j)}} \frac{1}{2} \sum_{k \in Q1} \log_2(1 + P_{k(j)}^{S1} a_{k(j)} + b_{k(j)}) \quad (15)$$

其中 $Q1$ 为源和中继在第二个传输阶段的子载波功率不同时为 0 的子载波集合,设 $Q1$ 元素数为 $N1_{(j)}$ 。用 Karush-Kuhn-Tuck(KKT) 条件对(15)式求解得源在第 1 阶段的最优功率分配为

$$P_{k(j)}^{S1} = \left(\frac{P^{S1}}{N1_{(j)}} + \frac{1}{N1_{(j)} \sum_{i \in Q1} \frac{1}{a_{i(j)}}} + \frac{1}{N1_{(j)} \sum_{i \in Q1} \frac{b_{i(j)}}{a_{i(j)}} - \frac{1}{a_{k(j)}} - \frac{b_{k(j)}}{a_{k(j)}}} \right)^+ \quad k \in Q1 \quad (16)$$

其中 $(x)^+ = \max(0, x)$ 表示取非负值。(16)式中功率为负的子载波将分配的功率置零,这将破坏节点总功率约束。为满足节点总功率约束并考虑算法实现复杂度,依据每个子载波的分配功率占总功率的比例对其功率进行调节,调节后的第 k 个子载波上的分配功率为

$$P_{k(j)}^{S1} = P_{k(j)}^{S1} P^{S1} / \left(\sum_{i=1}^N P_{i(j)}^{S1} \right) \quad (17)$$

第2步: $S-D$ 链路的注水功率分配

固定 $\mathbf{P}_{S1}^{(j)}, \mathbf{P}_R^{(j-1)}$, 最大化系统容量的功率分配

问题表示为

$$\begin{aligned} C_{(j)}^d &= \arg \max_{P_{S2}^{(j)}} \frac{1}{2} \sum_{k=1}^N \log_2 \\ &\left(1 + \left(\frac{P_{k(j)}^{S2} H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^{S2} H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0} \right) \right) \\ \text{s. t.} \quad \mathbf{1}^T \mathbf{P}_{S2}^{(j)} &= P^{S2} \quad \text{且} \quad \mathbf{P}_{S2}^{(j)} \geq 0 \end{aligned} \quad (18)$$

$$\begin{aligned} \text{令 } c_{k(j)} &= \frac{H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0}, \\ d_{k(j)} &= \frac{P_{k(j-1)}^R H_k^{RD} P_{k(j)}^{S1} H_k^{SR}}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0}, \end{aligned} \quad (18) \text{ 式的}$$

目标函数表示为

$$C_{(j)}^d = \arg \max_{P_{S2}^{(j)}} \frac{1}{2} \sum_{k=1}^N \log_2 (1 + P_{k(j)}^{S2} c_{k(j)} + d_{k(j)}) \quad (19)$$

使用 KKT 条件对(19)式求解可得第2阶段源节点在第 k 个子载波上的功率分配为

$$\begin{aligned} P_{k(j)}^{S2} &= \left(\frac{P^{S2}}{N} + \frac{1}{N} \sum_{i=1}^N \frac{1}{c_{i(j)}} \right. \\ &\left. + \frac{1}{N} \sum_{i=0}^N \frac{d_{i(j)}}{c_{i(j)}} - \frac{1}{c_{k(j)}} - \frac{d_{k(j)}}{c_{k(j)}} \right)^+ \end{aligned} \quad (20)$$

采用与第1步相同的成比例方法调节每个子载波上的功率,使满足第2阶段总功率约束

$$P_{k(j)}^{S2} = P_{k(j)}^{S2} P^{S2} / \left(\sum_{i=1}^N P_{i(j)}^{S2} \right) \quad (21)$$

第3步: $R-D$ 链路的注水功率分配

由于 $P_{k(j)}^R$ 同时存在于分子和分母,将分母项的 $P_{k(j)}^R$ 用 $P_{k(j-1)}^R$ 代替,并在注水功率分配时将 $P_{k(j-1)}^R$ 作为常数。固定 $\mathbf{P}_{S1}^{(j)}, \mathbf{P}_{S2}^{(j)}, \mathbf{P}_R^{(j-1)}$, 最大化系统容量的功率最优化的问题表示为

$$\begin{aligned} C_{(j)}^d &= \arg \max_{P_R^{(j)}} \frac{1}{2} \sum_{k=1}^N \log_2 \\ &\left(1 + \left(\frac{P_{k(j)}^{S2} H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^R H_k^{RD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^{S2} H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0} \right) \right) \\ \text{s. t.} \quad \mathbf{1}^T \mathbf{P}_R^{(j)} &= P^R \quad \text{且} \quad \mathbf{P}_R^{(j)} \geq 0 \end{aligned} \quad (22)$$

在 $R-D$ 链路功率分配时,若第 k 个子载波在第一步 $S-R$ 链路的功率分配 $P_{k(j)}^{S1} = 0$, 则中继节点在这个子载波上没有收到信息,因此设置中继节点在这个子载波上的发送功率 $P_{k(j)}^R = 0$ 。寻找 $S-R$ 链路所有功率分配为0的子载波,中继节点的

总功率在剩余子载波上分配,系统容量最大化的问题的目标函数表示为

$$\begin{aligned} C_{(j)}^d &= \arg \max_{P_R^{(j)}} \frac{1}{2} \sum_{k \in Q2} \log_2 \\ &\left(1 + \left(\frac{P_{k(j)}^{S2} H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^R H_k^{RD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^{S2} H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0} \right) \right) \end{aligned} \quad (23)$$

其中 $Q2$ 为 $S-R$ 链路分配功率不为0的子载波集合,假设此集合中元素的个数为 $N2_{(j)}$ 。

$$\begin{aligned} \text{令 } e_{k(j)} &= \frac{H_k^{RD} P_{k(j)}^{S1} H_k^{SR}}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0} f_{k(j)} \\ &= \frac{P_{k(j)}^{S2} H_k^{SD} P_{k(j)}^{S1} H_k^{SR} + P_{k(j)}^{S2} H_k^{SD} N_0}{(P_{k(j)}^{S1} H_k^{SR} + P_{k(j-1)}^R H_k^{RD} + N_0) N_0}, \end{aligned} \quad (23) \text{ 式目标函}$$

数就表示为

$$C_{(j)}^d = \arg \max_{P_R^{(j)}} \frac{1}{2} \sum_{k \in Q2} \log_2 (1 + P_{k(j)}^R e_{k(j)} + f_{k(j)}) \quad (24)$$

使用 KKT 条件对(24)式求解可得中继节点在第2阶段的功率分配为

$$\begin{aligned} P_{k(j)}^R &= \left(\frac{P^R}{N2_{(j)}} + \frac{1}{N2_{(j)}} \sum_{i \in Q2} \frac{1}{e_{i(j)}} \right. \\ &\left. + \frac{1}{N2_{(j)}} \sum_{i \in Q2} \frac{f_{i(j)}}{e_{i(j)}} - \frac{1}{e_{k(j)}} - \frac{f_{k(j)}}{e_{k(j)}} \right)^+ \quad k \in Q2 \end{aligned} \quad (25)$$

采用成比例的方法调节每个子载波的功率,使中继节点满足总功率的约束。

$$P_{k(j)}^R = P_{k(j)}^R P^R / \left(\sum_{i=1}^N P_{i(j)}^R \right) \quad (26)$$

为了保证及时退出迭代,设置一个常数 Δ ,每次迭代后由(3)式计算系统容量,如第 j 次迭代后的系统容量与第 $j-1$ 次迭代后的系统容量之差小于 Δ ,即 $C_{(j)}^d - C_{(j-1)}^d < \Delta$,停止迭代。

3 性能仿真

通过蒙特卡洛仿真验证 SIPAA 的性能。仿真中,每次随机产生 $S-R$ 、 $R-D$ 、 $S-D$ 链路 $N=16$ 的 Rayleigh 衰落子载波信道,令 $\alpha=2$,系统总发送功率 $P, N_0=1, \Delta=0.1$ 。 $S-D$ 距离 $d^{S,D}=1$ 。中继节点位于源与目的节点的连线上,距源节点距离 r ,距离目的节点距离 $1-r$ 。假设源节点和中继节点有相同的最大总发送功率。源节点的总发送功率需在第1阶段和第2阶段进行分配。第1个阶段源节点的最大发送功率为 $P^{S1}=a_1 P$,第2阶段的最大发送功率为 $P^{S2}=a_2 P$,其中 $a_1+a_2=0.5$,中继节点的最大发送功率为 $P^R=a_3 P$,其中 $a_3=0.5$ 。

图 1 所示为 $r = 0.5, a_1 = 0.25, a_2 = 0.25$ 时 SIPAA 和每个子载波上等功率分配算法的系统容量比较。可以观察到 SIPAA 的系统容量明显优于等功率分配的系统容量。在整个仿真区间里, SIPAA 相对于等功率分配有 1.5dB 的 SNR 节省。

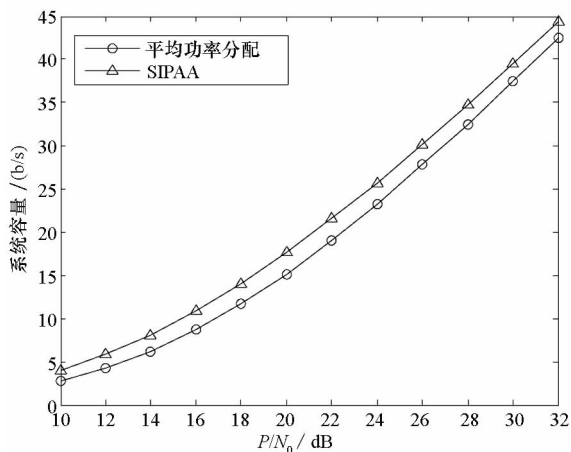


图 1 SIPAA 与等功率分配的系统容量比较

Fig. 1 System capacities of SIPAA and equal power allocation

图 2 所示为 $a_1 = 0.25, a_2 = 0.25, P/N_0 = 24\text{dB}$ 时 SIPAA 的平均迭代次数与中继节点位置的关系。可以观察到随着中继节点逐渐靠近目的节点, 平均迭代次数逐渐增加。文献[11]中指出 OFDM 系统, 发送 SNR 足够大时, 等功率分配的性能接近于注水功率分配的性能。由于源节点和中继节点的总发送功率相同, 那么第 2 阶段源节点的总发送功率小于中继节点的发送功率时, 中继传输的性能主要由 $S-R$ 链路决定, 随着 $S-R$ 距离的增加, 注水功率分配与等功率分配的性能差异也越大, 因此需要更多次的迭代。从图中可以观察到 SIPAA 需要的迭代次数很少, 即使 $r = 0.9$ 时, 也仅需要约 2.6 次的迭代。

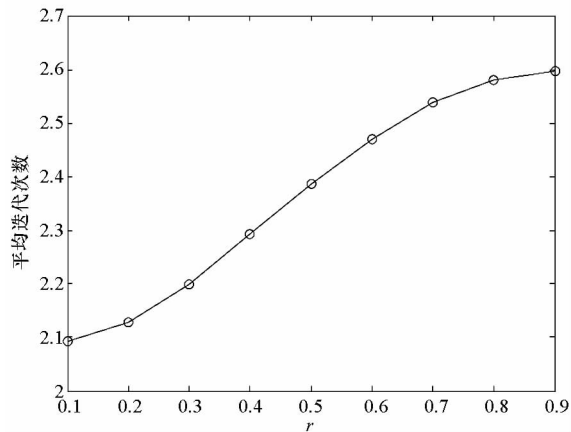


图 2 SIPAA 平均迭代次数与中继节点位置的关系

Fig. 2 The average times of iteration vs. the position of SIPAA the relay

图 3 所示为 a_1 分别为 0.1、0.25、0.4 时系统容量随中继节点的位置的关系, r 从 0.1 开始, 以步长 0.1 逐渐增加到 0.9。观察到在 $a_1 = 0.1$ 和 $a_1 = 0.25$ 时, 系统容量随着中继节点距离源节点的增加而减小。 $a_1 = 0.4, r = 0.1$ 时的系统容量达到最大。就是说在源节点和中继等功率分配时, 系统容量最大的位置在中间偏离源节点的位置, 这主要是由于第 2 阶段的传输是源节点和中继节点共同作用的效果, 同时源节点第 2 阶段占的功率越多, 则系统容量最大位置越靠近源节点。随 a_1 的增加, 系统容量变化的幅度在下降, 因此在中继节点靠近源节点时, a_1 越小, 系统容量越大, 随中继节点逐渐靠近目的节点, a_1 越大, 系统容量越大。

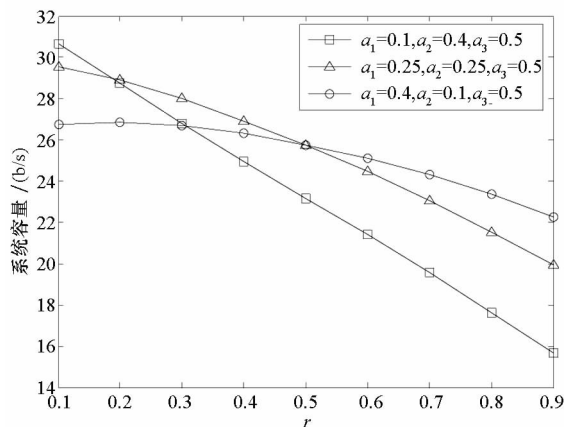


图 3 SIPAA 系统容量随中继节点位置的关系

Fig. 3 The average system capacity of SIPAA vs. the position of the relay

4 结束语

本文针对由源节点和单中继节点构成的协同 OFDM Alamouti STBC 系统的系统容量问题进行研究, 推导了系统容量的解析表达式。在节点功率分别受限的约束下, 提出了分步迭代算法来对每个节点注水功率分配。注水功率分配时由于进行了线性化处理, 具有较低的算法复杂度。仿真表明该算法所需迭代次数少, 并且相对于传统的等功率分配算法, 其系统容量可以获得显著的提升。

参考文献 (References)

- [1] Tarokh V, Jafarkhani H, Calderbank A. Space-time block coding from orthogonal designs [J]. IEEE Transactions Inform Theory, 1999, 45(5): 1456 - 1467.
- [2] 张天魁, 程婕, 冯春燕. 一种全速率的多天线中继 STBC 协作机制 [J]. 西安电子科技大学学报, 2010, 37(2): 334 - 339.

ZHANG Tiankui, CHENG Jie, FENG Chunyan. Full rate

STBC cooperation scheme for the multi-antenna relay [J]. Journal of Xidian University, 2010, 37(2): 334 – 339. (in Chinese)

[3] Laneman J, Wornell G. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks [J]. IEEE Transactions Inform Theory, 2003, 49(10): 2415 – 2425.

[4] Dharmawansa P, McKay M R, Mallik R K. Dual hop MIMO relaying with orthogonal space-time block codes [C]// IEEE International Conference on Communication 2009, Dresden, June 2009:1 – 5.

[5] He J, Kam P Y. Adaptive cooperative space-time block coding with amplify-and-forward strategy [C]// IEEE Vehicular Technology Conference Spring 2008, Singapore, May 2008: 1236 – 1240.

[6] He J, Kam P Y. Cooperative space-time block coding with amplify-and-forward strategy: exact bit error probability and adaptive forwarding schemes [J]. Physical Communication, 2008, 1(3): 209 – 220.

[7] Jang J, Lee K B, Lee Y H. Transmit power and bit allocations for OFDM systems in a fading channel [C]// IEEE GLOBECOM 2003, Dec 2003: 858 – 862.

[8] Li Y, Wang W B, Kong J, et al. Power allocation and subcarrier pairing in OFDM-based relaying system [C]// IEEE International Conference on Communication 2008, May 2008: 2602 – 2606.

[9] Shen Z H, Wang X X, Zhang H T. Power allocation and subcarrier pairing for OFDM based AF cooperative diversity systems [C]// IEEE Vehicular Technology Conference Spring 2009, Barcelona, April 2009: 1 – 5.

[10] Hammerstrom I, Wittneben A. On the optimal power allocation for Nonregenerative OFDM relay links [C]// IEEE International Conference on Communication 2006. Istanbul, June 2006: 4463 – 4468.

[11] Wang L J, Zhao H T, Tang Y, et al. The joint relay selection with power allocation for outage-optimization in OFDM-based cooperative relaying [C]// IEEE ICCASM 2010, Tai Yuan, China, Oct 2010: 557 – 561.

(上接第 124 页)

[3] 李金梁, 罗佳, 常宇亮, 等. 基于天线空域极化特性的虚拟极化接收技术[J]. 电波科学学报, 2009, 24(3): 389 – 393.

LI Jinliang, LUO Jia, CHANG Yuliang, et al. Virtual polarization receiver based on the spatial polarization characteristics of antenna [J]. Chinese Journal of Radio Science, 2009, 24(3): 389 – 393. (in Chinese)

[4] 罗佳, 王雪松, 李永祯, 等. 雷达目标极化散射矩阵测量的新方法研究[J]. 信号处理, 2009, 25(6): 868 – 873.

LUO Jia, WANG Xuesong, LI Yongzhen, et al. A novel polarimetric scattering matrix measurement method [J]. Signal Processing, 2009, 25(6): 868 – 873. (in Chinese)

[5] 戴幻尧, 李永祯, 王雪松, 等. 基于和差波束极化特性的目标极化散射矩阵测量方法研究[J]. 电子与信息学报, 2010, 32(4): 6 – 12.

DAI Huanyao, LI Yongzhen, WANG Xuesong, et al. A new target scattering matrix measurement algorithm based on polarization characteristics of sum-and-difference beam [J]. Journal of Electronics & Information Technology, 2010, 32(4): 6 – 12. (in Chinese)

[6] Dai H Y, Wang X S, Li Y Z, et al. Spatial polarization characteristics and scattering matrix measurement of orthogonal polarization binary array radar[J]. Science in China (Series F), 2010, 53(12): 2687 – 2695.

[7] 刘勇, 李永祯, 王雪松, 等. 有源压制干扰背景下的雷达目标空域极化检测方法[J]. 电波科学学报, 2010, 25(3): 445 – 450.

LIU Yong, LI Yongzhen, WANG Xuesong, et al. Radar target spatial polarization detection under active suppressive jamming background[J]. Chinese Journal of Radio Science, 2010, 25(3): 445 – 450. (in Chinese)

[8] 戴幻尧, 李永祯, 陈志杰, 等. 电扫偶极子相控阵天线的空域极化特性分析[J]. 国防科技大学学报, 2010, 32(1): 84 – 89.

DAI Huanyao, LI Yongzhen, CHEN Zhijie, et al. Spatial polarization characteristics of electronically scanning dipole phased arrays antenna [J]. Journal of National University of Defence Technology, 2010, 32(1): 84 – 89. (in Chinese)

[9] 张光义. 相控阵雷达系统[M]. 北京: 国防工业出版社, 1994.

ZHANG Guangyi. Phased-array radar system [M]. Beijing: Defense Industry Press, 1994. (in Chinese)

[10] Robert J M. Phased array antenna handbook [M]. 2nd ed. Artech House, 2005.

[11] Josefsson L, Persson P. Conformal array antenna theory and design[M]. IEEE Press, 2006.

[12] Ishimaru A. Wave propagation and scattering in random media [M]. IEEE Press, 1997.

[13] Ludwig A C. The definition of cross polarization[J]. IEEE Tran. on AP, 1973: 116 – 119.

高性能 RTI 自适应通信机制研究*

梁洪波^{1,2}, 柳林³, 姚益平¹, 刘步权¹, 慕晓冬², 李猛¹

(1. 国防科技大学 计算机学院, 湖南 长沙 410073;
2. 第二炮兵工程大学, 陕西 西安 710025; 3. 海军装备研究院, 北京 100161)

摘要: 为了提高 HLA 仿真系统在高性能计算平台上的运行性能, 针对高性能计算平台的高性能层次式通信环境, 提出了一种共享内存/IB/以太网自适应 RTI 通信机制——SACM 机制, 该机制能够根据通信双方所在节点的网络拓扑信息, 自适应建立延迟最小的通信连接, 实现共享内存、IB 等高速互连及以太网互连的自适应选择。同时在该通信机制中, 针对节点内高速共享内存通信需求, 提出了一种面向多盟员互连的共享内存通信算法, 有效提高了节点内盟员通信效率。对象属性值更新延迟测试结果表明, 该通信机制可以有效提高 RTI 在高性能计算平台的通信性能。

关键词: 分布式交互仿真; 运行支撑环境(RTI); 高性能计算平台; 自适应通信机制; 共享内存通信
中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0148-06

Research on self-adaptive communication mechanism
for high performance RTI

LIANG Hongbo^{1,2}, LIU Lin³, YAO Yiping¹, Liu Buquan¹, MU Xiaodong², Li Meng¹

(1. College of Computer, National University of Defense Technology, Changsha 410073, China;
2. Second Artillery Engineering University, Xi'an 710025, China;
3. Naval Academy of Armament, Beijing 100161, China)

Abstract: To promote the execution efficiency of HLA system which executes on high performance computing (HPC) platform, a self-adaptive communication mechanism, named SACM, is proposed for the hierarchy communication environment of HPC. By this mechanism, the connection with lowest latency can be established from the network information of computing nodes. The shared memory, high speed IB network and Ethernet network can be selected by federates adaptively. Aiming at the demand of shared memory communication between federates within one node, a high speed shared memory communication algorithm based on distributed infrastructure is proposed, and the communication performance can be improved by this algorithm. Experiment result of object attributes value updating latency has shown that the communication performance of RTI executing on HPC platform has been greatly improved.

Key words: distributed interactive simulation; runtime infrastructure; high performance computing platform; self-adaptive communication mechanism; shared memory communication

高层体系结构 HLA(High Level Architecture) 为分布式交互仿真提供了一个通用的技术框架和开放的标准^[1]。运行支撑环境(Runtime Infrastructure, RTI) 是 HLA 接口规范的具体实现^[2-3]。基于 HLA/RTI 的大规模仿真应用在运行过程中往往需要进行大量的通信, 如军事作战仿真中的雷达实体在每一步仿真过程中都需要接收其探测范围之内的导弹、飞机等众多实体的位置信息, 而当战场中雷达、导弹、飞机及舰船等作战实体的规模增大时, 系统的通信量往往成级数增长; 另一方面, 为保持系统因果序正确, 仿真应

用在运行过程中往往需要进行频繁的协同操作, 这种协同操作每次都需要交换大量的信息, 大量的通信已成为影响整个系统运行效率的主要因素之一^[4-5]。高性能计算机可为 HLA 仿真提供多种不同层次的高性能硬件通信环境^[6-7], 如一个节点内的共享内存、节点之间的 IB 高速互连、高性能节点与外部子系统的以太网互连等。然而, 目前的 HLA 运行支撑平台 RTI 大都只提供 TCP/IP 通信支持, 难以有效发挥高性能计算节点内共享内存通信及节点间 IB 通信的优势。因此, 研究面向高性能计算环境的 RTI 通信机制对于提高

* 收稿日期: 2011-09-13
基金项目: 国家自然科学基金资助项目(61170047)
作者简介: 梁洪波(1983—), 男, 河北邢台人, 博士研究生, E-mail: lianghongbo@nudt.edu.cn;
姚益平(通信作者), 男, 研究员, 博士, 博士生导师, E-mail: ypyao@nudt.edu.cn

HLA 仿真的运行性能和可扩展性等具有重要的理论和现实意义。

1 高性能计算机层次式通信环境及传统 RTI 通信机制的不足

高性能计算机(HPC)体系结构主要有对称多处理器共享存储并行机(SMP)、分布共享存储并行机(DSM)、大规模并行机(MPP)和 SMP 集群等^[8]。目前,以 SMP 集群(图 1)为代表的层次并行体系结构计算机发展迅速,已成为主流体系结构。这种结构的特点是结合了 SMP 和 MPP 的优点,同时具备节点内共享存储和节点间分布存储的层次结构,可实现节点内共享内存、节点间 Infiniband 等高速互连两级通信,此外,当与外部网络连接时,还存在高性能节点与外部 PC 节点间以太网通信。这种多层次通信环境的存在,客观上对基于高性能计算机的 HLA 仿真支撑平台的通信机制提出了更高的要求。

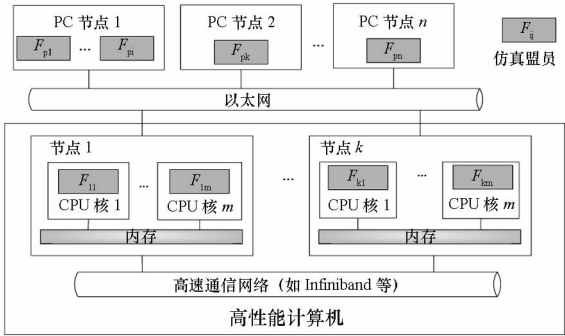


图 1 基于高性能计算平台的分布式仿真应用通信环境

Fig.1 Communication environment of distributed simulation application on HPC

随着仿真应用的不断发展,基于 PC 机网络的 HLA 仿真已越来越不能满足其对计算和通信能力的需求,基于高性能计算平台的 HLA 仿真正成为分布式仿真发展的一个重要趋势。在这类仿真应用中,HLA 联盟通常部署于图 1 所示的以高性能计算机为中心的网络环境中,高性能计算机通过其强大的计算能力与高速的通信能力为仿真应用的复杂模型解算、大规模实体交互等提供硬件支持^[9-10];而通过以太网与高性能计算机相连的 PC 机则为人机交互提供支持。高性能计算机的层次式通信结构可为盟员间通信提供多种可选的通信方式,例如在图 1 中,位于同一个高性能计算节点内的不同盟员进程间(F_{i1} 、 \cdots 、 F_{im}),可以通过共享内存、IB 或以太网完成盟员间通信;位于不同高性能计算节点间的盟员(如 F_{i1} 与

F_{kl})可以通过 IB 或以太网完成盟员间通信等。由于不同的通信方式性能差异很大,通常情况下,共享内存的通信性能优于 IB 网络,IB 网络优于以太网,如基于某高性能计算平台上的通信延迟测试结果:当数据包为 4KB 时,基于共享内存的进程间平均通信延迟为 $3.6\mu\text{s}$;基于 IB 的平均通信延迟为 $15\mu\text{s}$;基于千兆以太网的平均通信延迟为 $140\mu\text{s}$ 。因此,为了获得更高的仿真执行效率,面向高性能计算环境的 RTI 通信机制需要能够针对不同的环境选择最佳的通信方式,如节点内共享内存通信、节点间 IB 通信等^[11]。

然而,由于现有的 RTI 仿真平台主要面向分布式网络环境,盟员间通过以太网进行通信,受通信环境的制约,其通信机制主要基于 TCP/IP 实现,如 pRTI^[12]、KD-RTI^[13]、Portico RTI^[14]、ceRTI^[15]、StarLink^[16-18]等,虽然能满足基于分布网络构建 HLA 仿真系统的应用需求,但是,这种单一的通信机制不能有效发挥高性能计算机层次式通信硬件的效能,其主要原因有:(1)不能实现同一个计算节点内盟员间共享内存通信;(2)只能基于同一个网络实现盟员间互操作,不能发挥高性能计算节点间高速 IB 互连优势,例如使用该 RTI 基于图 1 的通信环境构建 HLA 仿真应用时,为了实现与外部盟员间通信,所有盟员必须基于以太网 TCP/IP 通信,不能基于 IB 网实现高性能计算节点间盟员高速通信。针对这些问题,文献^[11]给出了我们的高性能 RTI 的基本框架,提出在高性能 RTI 中兼容共享内存、IB 高速通信以及以太网通信等不同通信方式以提高通信性能,并从面向对象软件设计角度设计了基本实现框架。本文则对高性能 RTI 自适应通信机制进行深入研究,从自适应连接算法、共享内存通信算法等方面出发,研究如何尽可能减少通信延迟,以充分发挥高性能计算机的高速通信优势。

2 高效自适应 RTI 通信机制

针对单一的 TCP/IP 通信在高性能计算平台的不足,为了充分利用高性能计算机层次式通信环境的优势,论文提出了一种高效的共享内存/IB/以太网自适应 RTI 通信机制——SACM(Self-Adaptive Communication Mechanism),并针对 SACM 中的自适应连接与分布式仿真多盟员互连通信的特点,分别提出了基于位置和网络性能的自适应连接算法与基于分布式共享内存通信结构的高速内存通信算法。

2.1 SACM 设计

SACM 将共享内存通信与多网络 TCP/IP 通信相结合,使得通信双方盟员能够根据所在节点的网络信息,自适应选择最佳的通信方式:如果都位于同一个计算节点内部,则选择共享内存实现盟员间通信;如果分别位于不同的高性能计算节点,则选择 IB 高速互连网络实现盟员间通信;如果分别位于高性能计算节点与 PC 节点,或者分别位于不同的 PC 节点内,则选择以太网网络实现盟员间通信。SACM 主要包括通信初始化、连接建立、消息发送及断开连接 4 部分,如图 2 所示。

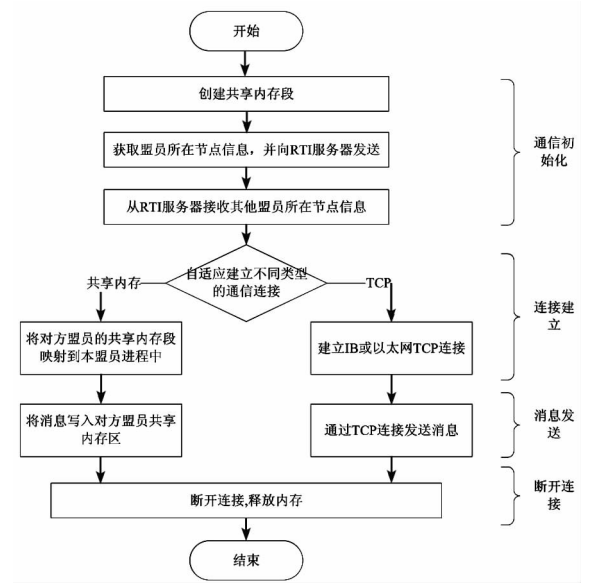


图 2 SACM 通信过程

Fig. 2 Communication process of SACM

(1)通信初始化。盟员进程在加入联盟执行时,首先创建用于共享内存通信的接收内存段。然后获取本盟员进程所在节点的网络信息(包括 IP 地址,网络带宽等),将信息提交到 RTI 服务器,并从服务器接收其他已加入盟员所在节点信息,使得每个盟员方都拥有全局的网络信息。

(2)通信连接建立。在获取了全局的网络信息后,盟员将根据通信双方进程所在节点的网络信息,自适应建立延迟最小的通信连接。如果需要建立共享内存通信连接,则通信双方互相映射对方的共享内存区到本进程中;否则建立 IB 或以太网 TCP 通信连接,建立连接后,在盟员方记录连接类型及连接指针。

(3)消息发送。当通信双方需要发送消息时,需要根据记录的连接类型选择不同的发送方式:如果为共享内存通信,将消息写入对方盟员的共享内存区;否则,获取 TCP 连接指针,通过 TCP 连接发送消息。

(4)通信连接断开。在盟员退出联盟执行时,该盟员与所有其他盟员间 TCP 通信连接将断开,然后该盟员释放其创建的共享内存段,通信结束。

SACM 中,通信初始化以及连接建立过程与基于 TCP/IP 的通信机制相比,需要一定的额外开销,例如与 RTI 服务器的网络信息同步,共享内存段的创建与映射、以及根据不同网络类型建立不同的 TCP 连接等。为了减少这些额外开销对仿真系统性能的影响,在实现中,通信初始化以及连接建立过程在盟员加入联盟执行服务调用中完成,在整个仿真过程中,每一个盟员只执行一次该服务调用。同时,由于仿真系统运行性能主要取决于仿真推进速度,为了减少仿真推进过程中自适应选择引起的消息发送开销,不同通信方式的自适应选择通过简单的标志位(该标志位在连接建立过程中记录)判断即可,与基于 TCP/IP 的通信机制相比,在仿真推进过程中几乎不会产生额外的开销。

2.2 基于位置和网络性能的自适应连接算法

SACM 的关键是盟员间如何自适应建立最佳的通信连接,针对该问题,提出了基于位置和网络性能的自适应连接算法,该算法基于如下原理:在盟员的 IP 地址列表(位置信息)中,位于同一个节点内的盟员必然有一个 IP 地址相同,此时为该盟员间通信指定共享内存通信方式;否则指定通信方式为 TCP 通信。由于节点间 IB 高速网络的带宽(可达 40Gb/s)明显大于外部以太网(千兆网带宽 1Gb/s)。因此可以根据网络带宽信息区分高速与低速网络,从而优先选择高带宽网络建立 TCP 连接,这样既保证了高性能节点间盟员通信可以基于高速互连网络实现,又满足了高性能计算机内部与外部盟员间通信需求。同时,由于两个盟员间可以共享同一 TCP 连接,这里规定由句柄值小的盟员发起连接。为了实现该算法,在盟员方定义了节点列表 NodeList(NL),用于按网络性能由好到差顺序存储盟员的节点信息(包括 IP 地址与对应的网络带宽),算法描述如下:

```
for(本地盟员  $F_m$  地址列表  $NL_m$  中每一个地址  $a_i$ )
{
    for(目标盟员  $F_n$  地址列表  $NL_n$  中每一个地址  $a_j$ )
    {
        if( $a_i = a_j$ )
             $F_m$  和  $F_n$  位于同一个节点上,选择共享内存通信方式,获取  $F_n$  共享内存指针,算法结束;
        else
             $F_m$  和  $F_n$  位于不同节点上,选择 TCP 通信方式;
```

```
if( $a_i$  与  $a_j$  在同一网段内)
    if( $F_m$  盟员句柄  $H_m$  小于  $F_n$  盟员句柄  $H_n$ )
        由  $F_m$  创建指向  $a_j$  的 TCP 连接;
    算法结束;
endif
endelse
```

盟员在加入联盟执行时,请求加入盟员和已加入盟员将根据盟员间通信需求,基于以上算法自适应建立盟员间通信连接,如图 3 所示。(1)盟员 A 与盟员 B 位于同一个高性能计算节点内,盟员间采用共享内存通信;(2)盟员 A 与盟员 C 位于不同的高性能计算节点,且同时通过 IB 与以太网互连,由于 IB 带宽高于以太网,因此基于 IB 网络建立 TCP 连接;(3)盟员 A 与盟员 D 分别位于高性能计算节点与 PC 节点,基于以太网建立 TCP 连接。

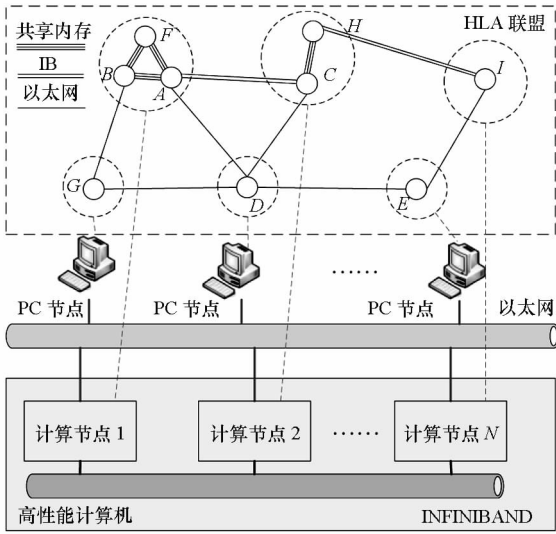


图3 基于 SACM 的 HLA 联盟及盟员间通信连接

Fig.3 HLA Federation based on SACM and connections between federates

2.3 基于分布式共享内存结构的高速通信算法

在基于高性能计算环境的 HLA 仿真中,往往有多个盟员并行运行在同一个高性能计算节点内,盟员之间存在大量而频繁的数据交互。在这样的通信环境中,如何管理共享内存并保证进程间内存访问的一致性共享内存通信正确性的关键,也是影响通信性能的最主要因素^[19]。针对该问题,本文提出了一种基于分布式共享内存结构的通信算法:每一个盟员维护一个属于本地盟员的共享内存段,用来接收其他盟员的消息。发送盟员通过向接收盟员的共享内存段写数据实现消息发送,目标盟员通过从本盟员的共享内存段读

数据实现消息接收(如图 4 所示)。由于每一个接收盟员都拥有独立的内存区域,因此具有不同目的地址的发送进程不会竞争同一个共享内存段,有效减少了盟员间共享内存访问冲突。同时,由于只需要保证局部共享内存的一致性,不需要全局进程同步,使得进程间的同步开销减小,保证了通信效率。

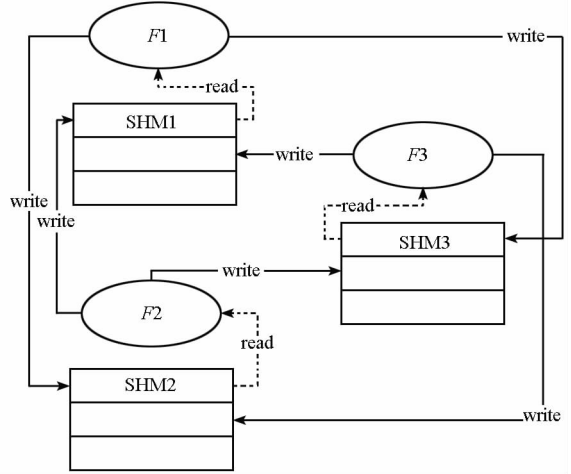


图4 分布式共享内存通信结构

Fig.4 Distributed shared memory communication infrastructure

内存中,消息分为两部分存储:消息长度与消息内容。消息长度占用 4 个字节内存空间,消息内容占用大小为消息长度字节的内存空间,进程间同步基于信号量机制实现,该算法由写算法和读算法组成,写算法描述如下:

for (每一个长度为 l 的消息 m)

- ① 等待写信号量 $write$;
- ② 比较可用空闲内存大小 a 与 l , if ($a < l + 4$) then 等待固定时间间隔 Δt , 转②; else 转③;
- ③ 比较写指针 w 与内存结束位置 e , if ($w + 4 \leq e$) then 将长度值 l 写入内存, $w = w + 4$, 转⑤; else 转④;
- ④ 将长度值分两部分存储,首先,取长度为 $e - w$ 的第一部分存储在 w 与 e 之间的内存区域,并移动 w 到内存起始位置 b ; 然后将剩余部分写入内存,移动 w 到长度值存储区的末尾;
- ⑤ if ($w + l \leq e$) then 将消息内容写入内存区,移动写指针 w , $w = w + l$, 转⑦; else 转⑥;
- ⑥ 将消息内容分两部分存储,第一部分存储在 w 与 e 之间的内存区,移动 w 到内存起始位置 b ; 再将第二部分写入 b 指向的内存区,移动 w 到消息内容末尾;
- ⑦ 等待互斥信号量 $compute$;
- ⑧ 计算可用空闲内存大小 a , $a = a - l - 4$;
- ⑨ 释放 $compute$ 信号量;
- ⑩ 释放 $full$ 信号量;

⑪ 释放 write 信号量;

}

读算法描述如下:

while(1)

{

① 等待信号量 full;

② 比较读指针 r 与内存结束位置 e , if($r+4 \leq e$) then 读出消息长度值 l , $r=r+4$, 转④; else 转③;

③ 将长度值 l 分两部分读出, 首先读出 r 与 e 之间的数据, 移动 r 到内存起始位置 b ; 再读出 $4-(e-r)$ 长度数据。两部分数据合并即得到长度值 l , 移动读指针 r 到长度值存储区的末尾;

④ if($r+l \leq e$) then 读出消息内容, 移动读指针 $r, r=r+l$, 转⑥; else 转⑤;

⑤ 将消息内容分两部分读出, 首先读出 r 与 e 之间的数据, 移动 r 到内存起始位置 b ; 再读出 $l-(e-r)$ 长度数据, 两部分数据合并即得到消息内容。移动读指针 r , 使 r 指向内存中该消息的结束位置;

⑥ 等待互斥信号量 compute;

⑦ 计算可用空闲内存大小 $a, a=a+l+4$;

⑧ 释放 compute 信号量;

}

3 性能测试

3.1 实验环境

本通信机制在课题组设计的高性能 RTI 软件 shmRTI 中得到实现。在 HLA/RTI 中, 对象属性值更新是 HLA 仿真系统中最常用的服务^[20], 当不使用时间管理时, 其性能直接反映 RTI 的通信性能。实验环境由 3 个高性能计算节点和 3 台 PC 机组成。高性能计算节点环境为双路四核 QuadCore Xeon 处理器, 2.53GHz, 32GB RAM, PC 机配置为 Intel 双核 E7500 处理器, 2.93GHz, 2GB RAM, 高性能节点之间在集群内部通过 IB 互连, 同时每一个高性能节点与 PC 机采用千兆以太网互连, 测试结果与目前国际上普遍使用的基于 TCP/IP 通信机制的 RTI 软件 pRTI 进行了对比, pRTI 作为一款通用的 RTI 软件, 具有较高的性能和较好的稳定性^[4], 在国内外各类 HLA 仿真系统中得到了广泛应用。

3.2 性能测试与分析

对象属性值更新延迟测试过程为: (1) 首先构造测试联盟 F , 联盟 F 由盟员 A, B, C, \dots, I 这 9 个盟员组成, 盟员在高性能计算机及 PC 机上的分布如图 3 所示。(2) 在联盟 F 中分别测试同一高性能节点内、高性能节点间、高性能节点与 PC

节点间的盟员对象属性值更新延迟。(3) 盟员 F_i 与 F_j 间对象属性值更新延迟测试方法如下: F_i 与 F_j 采用 Ping-Pong 方式发送属性更新, 即 F_i 首先发出属性更新, F_j 在接收到属性更新后, 返回一个新的属性更新, F_i 在收到后同样再发出一个新的属性更新, 如此循环 10 万次, 记录完成时间, 再重复 10 次, 计算平均完成时间 t (去除最长时间与最短时间), 根据 t 计算出平均端对端延迟 $l(l = [t / (100000 \times 2)] \times 1000000$, 单位: μs)。改变更新对象属性值的数据大小, 重复以上实验, 记录不同数据大小下的对象属性值更新延迟。

高性能节点内盟员通信延迟如图 5 所示, 高性能节点间盟员通信延迟如图 6 所示, 高性能节点与外部节点盟员通信延迟如图 7 所示。

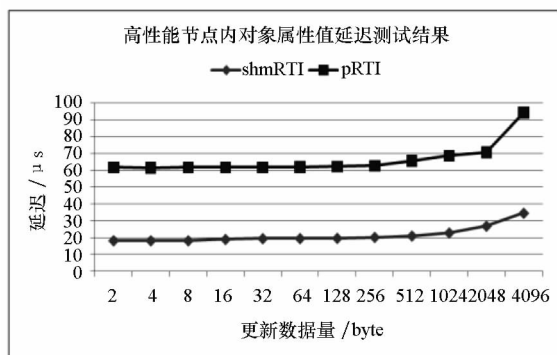


图 5 高性能节点内对象属性值更新延迟测试

Fig. 5 Object attributes value updating latency of federates on the same HPC node

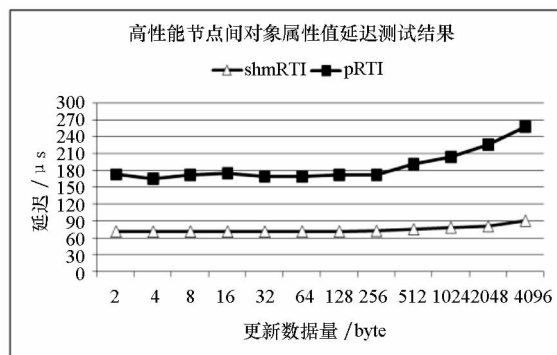


图 6 高性能节点间对象属性值更新延迟测试

Fig. 6 Object attributes value updating latency of federates on different HPC nodes

对象属性值更新延迟测试结果表明, 在同一个高性能计算节点内, shmRTI 能够自适应选择共享内存实现盟员间通信, TCP 虽然针对节点内通信做了优化, 但 pRTI 的延迟仍为 shmRTI 的 3 倍 (图 5); 在不同高性能计算节点间 (图 6), shmRTI 能够在高速 IB 网络与千兆网之间自动选择 IB 建立 TCP 连接, 而 pRTI 中, 所有盟员都必须位于同一个网段内, 因此, 为了兼顾与节点外的盟员通

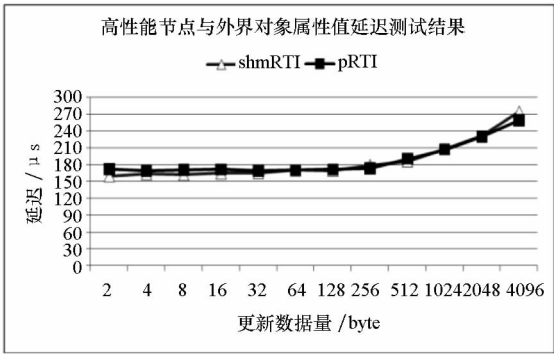


图 7 高性能节点与外界对象属性值更新延迟测试

Fig. 7 Object attributes value updating latency of federates between HPC node and outside node

信,高性能盟员间通信也必须通过千兆网实现,此时,shmRTI 的延迟约为 pRTI 的 1/3 ~ 2/5;而对于高性能盟员与外界盟员间通信(图 7),shmRTI 与 pRTI 都采用基于千兆网的 TCP 通信,其对象属性值更新延迟相当。测试结果充分说明了自适应通信机制更加适合于高性能计算平台。

4 结 论

论文针对基于 HPC 环境的大规模分布式交互仿真应用特点,分析了目前 RTI 通信机制在 HPC 环境中应用的不足之处,提出了一种共享内存/IB/以太网自适应 RTI 通信机制。该通信机制的创新在于通信进程能够根据所在节点的网络信息,自适应在共享内存、高速网络、以太网网络中做出选择,建立延迟最小的通信连接;同时通过高效的共享内存算法设计,有效提高了节点内盟员间的通信性能。性能测试表明,在高性能计算平台上,基于自适应通信机制的 RTI 通信性能明显优于基于 TCP/IP 通信机制的 RTI。

参考文献 (References)

[1] IEEE Standard for Modeling and Simulation. High Level Architecture [HLA] -Federate Interface Specification [S]. IEEE Std 1516.1 -2000, 2001.

[2] Li Z X. A Three-phases byzantine fault tolerance mechanism for HLA-based simulation. in distributed simulation and real time applications (DS-RT) [C]// 2010 IEEE/ACM 14th International Symposium, 2010.

[3] Fujimoto. Parallel and distributed simulation systems [M]. Wiley Interscience,2000.

[4] 姚益平. 高性能分布式交互仿真运行支撑平台关键技术研究[D]. 长沙:国防科学技术大学, 2003.

YAO Yiping. Research on technologies of high performance distributed simulation runtime infrastructure [D]. Changsha: National University of Defense Technology, 2003. (in Chinese)

[5] 张颖星. 面向复杂系统应用的并行离散事件仿真性能优化技术研究[D]. 长沙:国防科学技术大学, 2011.

ZHANG Yingxing. Research on performance optimization of

parallel discrete event simulation for complex system application [D]. Changsha: National University of Defense Technology, 2011. (in Chinese)

[6] Fujimoto. Design of high performance RTI software. in distributed simulation and real-time applications (DS-RT 2000) [C]// Proceedings of 4th IEEE International Workshop, 2000.

[7] Steinman J S, Blank G E, Brutocao J S. Design of the HPC-RTI for the high level architecture [C]// 99F-SIW - 067, 1999.

[8] 单莹, 吴建平, 王正华. 基于 SMP 集群的多层次并行编程模型与并行优化技术[J]. 计算机应用研究, 2006. 23 (10): 254 - 256.

SHAN Ying, WU Jianping, WANG Zhenghua. Hierarchical parallel programming model and parallelization and optimization techniques based on SMP cluster[J]. Application Research of Computers, 2006. 23 (10): 254 - 256. (in Chinese)

[9] 陈春鹏. 基于 HPC 环境的 RTI 及作战仿真应用[J]. 系统仿真学报, 2006, 18(10): 2794 - 2797.

CHEN Chunpeng. HPC-based RTI and martial simulation application[J]. Journal of System Simulation, 2006, 18(10): 2794 - 2797. (in Chinese)

[10] 姚益平, 张颖星. 基于并行处理的分析仿真解决方案[J]. 系统仿真学报, 2008. 20(24): 1375 - 1381.

YAO Yiping, ZHANG Yingxing. Solution for analytic simulation based on parallel processing [J]. Journal of System Simulation, 2008, 20(24): 1375 - 1381.

[11] Liang H B, Yao Y P, Mu X D. Development of a HLA/RTI for high performance simulation [J]. Advanced Materials Research, 2011, (291 - 294): 3044 - 3049.

[12] pRTI. Available from: <http://www.pitch.se/>.

[13] 唐见兵, 李革, 杨海刚. KD-RTI 软件的测试研究[J]. 计算机工程, 2007, 33(3): 62 - 64.

TANG Jianbing, LI Ge, YANG Haigang. Research on the test of KD-RTI software [J]. Computer Engineering, 2007, 33 (3): 62 - 64. (in Chinese)

[14] Portico website. Available from: www.porticoproject.org.

[15] Noulard E, Rousselot J Y. CERTI, An open source RTI, why and how[R]. 2009.

[16] Liu B Q. An efficient algorithm in the HLA time management [C]// Simulation Conference, 2006. WSC 06. Proceedings of the Winter, 2006.

[17] Liu B Q. Development of a runtime infrastructure for large-scale distributed simulations [C]// Simulation Conference, 2006. WSC 06. Proceedings of the Winter, 2006.

[18] 姚益平, 卢锡城. HLA/RTI 时间管理的一种优化算法[J]. 电子学报, 2004, 32(2): 294 - 297.

YAO Yiping, LU Xicheng. An optimized algorithm of HLA/RTI time management[J]. Acta Electronica Sinica, 2004, 32 (2): 294 - 297.

[19] Adelantado M. HP-CERTI: Towards a high performance, high availability open source RTI for composable simulations[C] // 04F-SIW-014, Orlando, United States, 2004.

[20] 张卫, 张童. 基于 HLA 分布式仿真的 Web 服务化[J]. 国防科技大学学报, 2008, 30(5): 120 - 124.

ZHANG Wei, ZHANG Tong. Web service enabling of HLA-based distributed simulation [J]. Journal of National University of Defense Technology, 2008, 30(5): 120 - 124.

原核细胞仿生自修复电路设计*

李岳,王南天,钱彦岭

(国防科技大学 装备综合保障技术重点实验室,湖南 长沙 410073)

摘要: 仿生电子阵列是模拟生物体自修复机制,具有在线自修复能力的新型电路。针对基于真核细胞的仿生电子阵列配置存储消耗大等不足,在比较真核细胞、原核细胞基因结构的基础上,提出了一种基于原核细胞的仿生电子阵列结构,设计实现了差分二进制相移键控调制电路,通过仿真验证了该结构的有效性。

关键词: 电路设计;原核细胞仿生;自修复;差分二进制相移键控调制电路

中图分类号: TP330 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0154-04

Self-healing circuit design inspired by prokaryotic cell

LI Yue, WANG Nantian, QIAN Yanling

(Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China)

Abstract: Bio-inspired array is a new type of circuit based on the self-healing mechanism of organism. The current study presents a novel prokaryotic cell-based bio-inspired array structure based on the gene structure of eukaryotic cell and prokaryotic cell. Simulation experiment on designing a self-healing DBPSK circuit is carried out and the result shows that the presented prokaryotic cell-based bio-inspired array structure is effective and consumes less memory storage than the traditional eukaryotic array structure.

Key words: circuit design; prokaryotic cell-based bio-inspired; self-healing; Differentially Binary Phase Shift Keying (DBPSK)

提高重要装备中电子系统的可靠性,一般通过关键部件完全备份(如三模冗余^[1])来实现,这种做法往往存在环境自适应能力差的问题^[2]。仿生电子阵列是近年来提出的一种具有类似生物体自诊断、自重构和自修复能力的新型电路,具有高可靠性,良好的环境适应性等优点,在人力不能直接维修或要求高可靠性的航空、航天机载电子设备等领域具有良好的应用前景^[2-4]。

仿生电子阵列一般是基于真核细胞的仿生,模拟多细胞生物组织发育过程中体现的细胞自复制、自修复能力。这种细胞阵列中,每个细胞都包含有整个阵列的全部配置信息,资源消耗大。本文利用原核细胞结构简单、遗传物质分散的特点,提出并实现了一种新型的细胞电路和阵列结构,拓展了仿生电子阵列的范围,大幅度减少了仿生电子阵列对存储器的消耗,对将仿生电子阵列推向应用,具有一定的意义与价值。

1 仿生电子阵列的基本原理

在生物学中,根据细胞在进化中的地位、结构复杂程度可将细胞分为原核细胞和真核细胞^[5]。

目前,一般的仿生电子阵列都是基于真核细胞的仿生,模拟胚胎的发育过程,故也常称为胚胎电子阵列,最开始是由 Mange 等人提出^[6],其基本原理如图 1 所示^[7]。

从阵列结构上看,胚胎电子阵列是由阵列细胞组成的均匀二维阵列,每个细胞的硬件结构完全相同,通过冯·诺依曼近邻连接^[8]形成二维阵列。从细胞内部结构看,细胞内包含一个配置存储模块,用来模拟真核生物细胞的 DNA,存储阵列需实现的功能。类似真核细胞的分裂与分化,电子细胞实现的具体功能由其根据自身位置选择配置存储模块中的一个配置来确定。

胚胎电子阵列中每个电子细胞都包含有整个阵列的全部配置信息,根据常用的行(列)或单细胞移出自修复机制^[7],细胞内的大部分配置信息将不会被使用,存储消耗大。

和真核细胞相比,原核细胞基因少、结构简单,但有形成相互合作群落的能力,如链球菌、四联球菌、八叠球菌、葡萄球菌等^[9]。基于此,本文借鉴原核细胞的结构特征,设计仿生电子阵列。

* 收稿日期:2011-12-06
基金项目:国家部委资助项目
作者简介:李岳(1964—),男,湖南长沙人,教授,博士,博士生导师,E-mail:liyue@nudt.edu.cn

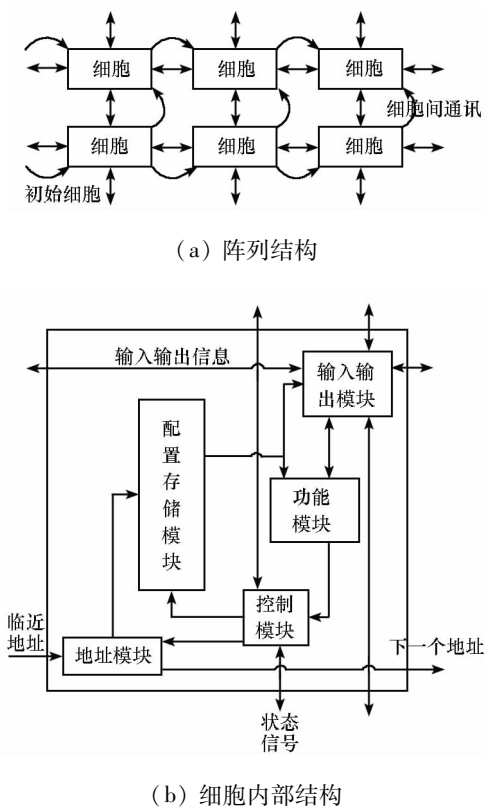


图1 仿生电子阵列基本结构

Fig.1 Basic structure of bio-inspired array

2 原核细胞仿生电路设计

2.1 细胞配置结构设计

真核细胞包含整个生物体的全部基因,不同细胞根据自身环境表达不同的基因而体现出不同的形态。和真核细胞类似,图1所示基于真核细胞仿生电子阵列的细胞结构中,其配置存储模块包含有整个阵列所有细胞的配置信息,每个细胞根据自身地址选择不同配置信息配置细胞以实现不同功能,其基本原理如图2所示。例如,某细胞的配置存储模块输入地址为 k ,则配置存储模块输出细胞 k 配置作为细胞的自身配置,完成该细胞的功能分化。

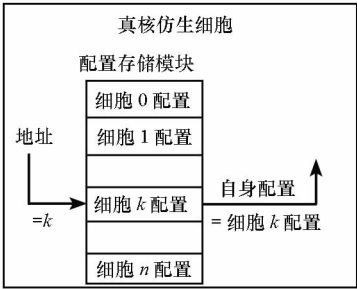


图2 真核仿生细胞配置

Fig.2 Cell configuration of eukaryotic cell inspired

原核细胞内遗传物质则比较少,除了自身生长所需的基因外,仅包含少量用于和周围细胞进

行信息交换等其他基因。对比原核细胞与真核细胞的基因特点,设计具有与原核细胞基因类似配置信息的电子细胞,其示意图如图3所示。

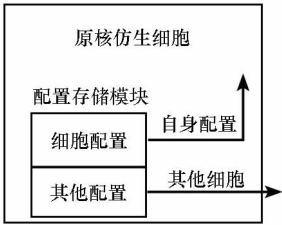


图3 原核仿生细胞配置

Fig.3 Cell configuration of prokaryotic cell inspired

将图3所示的仿原核电子细胞结构与图2所示的仿真核细胞结构相比,容易看出仿原核电子细胞配置信息要少很多,特别是当阵列细胞数量比较大的时候。假设仿原核电子细胞的其他配置和自身细胞配置数据量相等,如果阵列有10个细胞,则仿真核细胞阵列的配置信息是仿原核细胞阵列的5倍,如果有100个细胞,则是50倍。

2.2 阵列设计

如图1所示胚胎电子阵列中,细胞成二维均匀分布,当某细胞故障需要移出时,一般采用整行(列)移出的方式,许多正常细胞被移出。肺炎链球菌分裂沿平面进行,分裂后细胞成链状,借鉴此结构,本设计将阵列设计为链状的一维结构,细胞间依次连接,如图4所示。

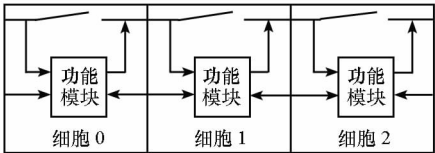


图4 基于原核细胞仿生的链状阵列

Fig.4 Catenulate array based on prokaryotic cell

另外,设计了可切断的总线结构,既可以满足电路中对远距离信号的传送要求,又实现了总线复用。例如需传递细胞0信号到细胞2,有两种方式。方式1:通过细胞0、1,细胞1、2之间的局部连接经由细胞1传递。方式2:细胞0的总线开关断开,细胞1的总线开关闭合并关闭总线驱动,就可以将信号传到细胞2。如果该信号还要传递给后面的细胞,将细胞2的开关闭合即可。由于细胞0的开关断开,其左边总线还可以用来接收其左侧细胞传来的信号,实现长总线的分段使用。

2.3 配置与重构

使用仿生设计的目的是实现电路的自修复,以提高可靠性,故要研究其自修复机制。

在如图 1 所示的细胞阵列结构中,当某细胞故障后移出,其右侧所有细胞依次完成自身左侧细胞的功能。故在如图 3 所示结构中,在“其他配置”中存储其左侧细胞的配置信息(故下文称“其他配置”为“备份配置”),当细胞故障时发出故障信号,触发其右侧细胞重构,如图 5 所示。

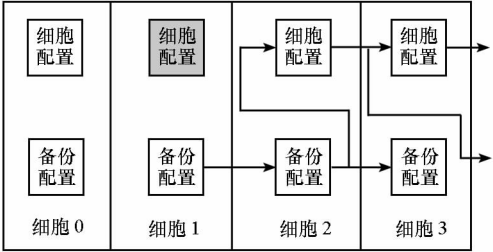


图 5 阵列重构原理

Fig. 5 Reconfiguration principle of the array

某时刻细胞 1 检测到细胞故障,向右发出故障信号。细胞 2 检测到细胞 1 发出的故障信号后,立刻将此信号向后传递给细胞 3,同时开始将其“备份配置”导入其“细胞配置”,并将其“细胞配置”向细胞 3 传递。细胞 3 检测到故障信号后,立刻接收细胞 2 传来的配置信息。故障细胞 1 左侧的细胞 0 则保持不变。

假想阵列最左侧细胞 0 的左侧还有一个虚拟的细胞,让该细胞发出故障信号,就可以利用细胞的自修复机制完成整个阵列的初始化,基本原理如图 6 所示。

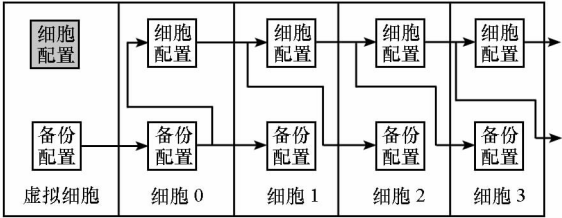


图 6 阵列初始化原理

Fig. 6 Initialization principle of the array

3 仿真实验验证

本实验实现一个通信领域的基本模块,差分二进制相移键控调制电路(Differentially Binary Phase Shift Keying, DBPSK)。根据 DBPSK 的基本原理^[10],用 Verilog HDL 描述后进行综合,得到如图 7 所示的电路。实现该电路只需要上节中设计的 3 个细胞,这里设计 6 个细胞的链状阵列。将细胞阵列用 Verilog HDL 实现,通过 ISIM 仿真,结果如图 8 所示。

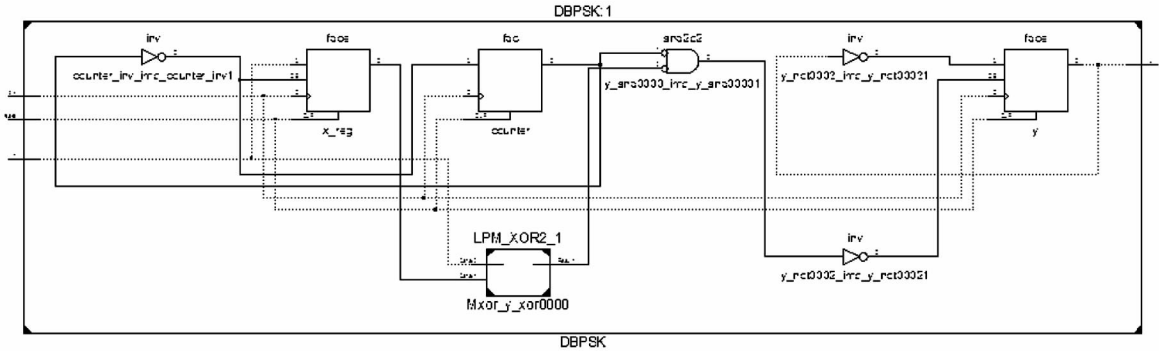


图 7 DBPSK 实现电路

Fig. 7 Realization circuit of DBPSK

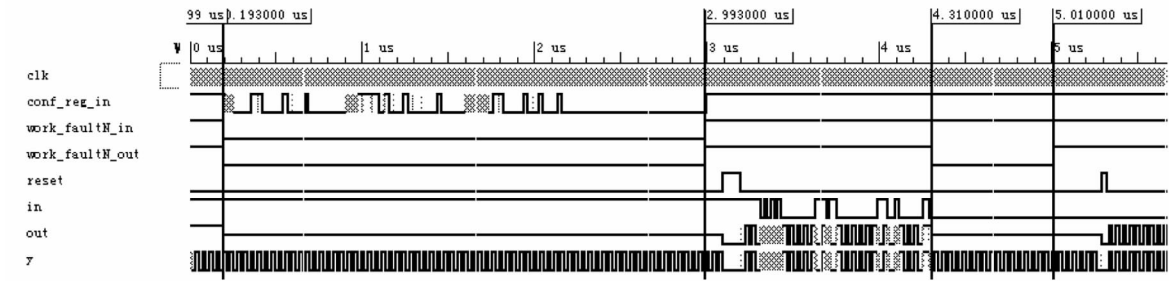


图 8 仿真结果

Fig. 8 Result of simulation

图 8 中,clk 为输入时钟,conf_reg_in 为配置输入信息。work_faultN_in 为工作/配置控制输入,work_faultN_out 为阵列的工作/配置状态输

出,低电平表示阵列处于配置(或重构)状态,高电平则表示工作状态。信号 reset、in、out 分别为细胞阵列实现 DBPSK 的重启、调制输入和调制输

出。信号 y 为直接用 DBPSK 模块的输出。

在 $work_faultN_in$ 的第 1 个下降沿(图中竖线所示的 $0.193\mu s$ 时刻), $work_faultN_in$ 变为低电平,阵列开始配置,载入 $conf_reg_in$ 的配置数据, $work_faultN_in$ 变为高电平($2.993\mu s$ 时刻)后,阵列开始工作。由于 $work_faultN_in$ 的配置信号会依次向右传播,故最后一个细胞 $work_faultN_out$ 输出有一个同样的低电平($0.193 \sim 2.993\mu s$)。 $work_faultN_out$ 的第二个下降沿($4.31\mu s$)的出现是因为内部某细胞故障,产生的重构信号依次向右传递,最后从 $work_faultN_out$

输出。

如图 9 所示, $work_faultN_out$ 为 0 时,阵列处于配置(或重构)状态, out 信号高阻输出(为 z ,如 $\sim 2.993\mu s$)。每次配置(或重构)后,由于阵列内部信号未知, out 输出不确定(为 x ,如 $2.993 \sim 3.093\mu s$),在输入 $reset$ 信号($3.093 \sim 3.193\mu s$)重启后, out 输出正常($3.193\mu s \sim$)。

通过观察图 9 中 in 、 out 信号以及比较 out 、 y 信号,容易看出该阵列实现了 DBPSK 的功能,且具有一定的自修复能力。

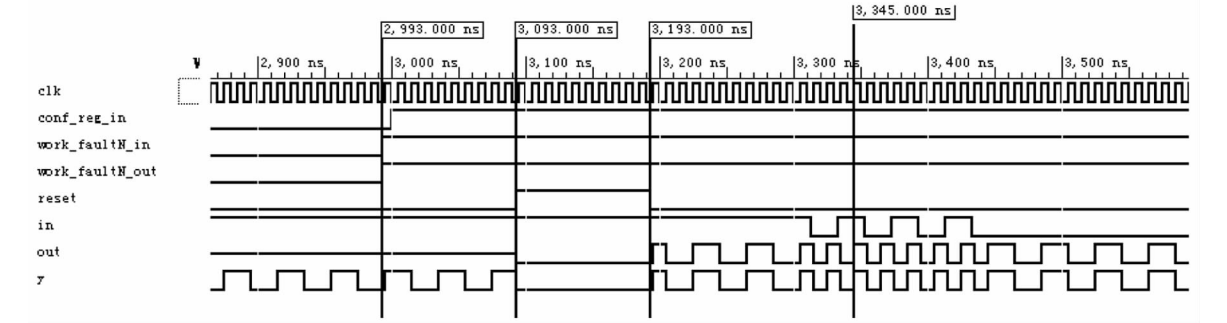


图 9 阵列调制结果
Fig. 9 Result of the array's modulation

4 结论与展望

本文以 DBPSK 为对象,提出并实现了一种新型的细胞电路和阵列结构,拓展了仿生电子阵列的范围。仿生电子阵列改变了传统电子系统的电路结构、容错模式以及数字电路系统的设计方法,在航空、航天等重要电子装备(系统)中具有广阔的应用前景。

参考文献 (References)

[1] Kim H, Jeon H J, Lee K. The design and evaluation of all voting triple modular redundancy system[C]// Proceedings of the 2002 Reliability and Maintainability Symposium, 2002: 439 - 444.

[2] 林勇. 基于进化型硬件的容错方法研究[D]. 合肥: 中国科学技术大学, 2007.

LIN Yong. Research fault tolerance methods based on evolvable hardware[D]. Hefei: University of Science and Technology of China, 2007. (in Chinese)

[3] 郑瑞娟, 王慧强. 生物启发的容错计算技术研究[J]. 计算机工程与应用, 2006, 42(4): 30 - 34.

ZHENG Ruijuan, WANG Huiqiang. A survey of bio-inspired fault tolerance computing techniques [J]. Computer Engineering and Technology, 2006, 42(4): 30 - 34. (in Chinese)

[4] 周贵峰, 钱彦岭, 王南天, 等. 基于胚胎型仿生硬件结构 FIR 滤波器设计与仿真[J]. 电子测量与仪器学报, 2010, 24: 61 - 65.

ZHOU Guifeng, QIAN Yanling, WANG Nantian, et al. Design and simulation of FIR filters based on embryonic bio-inspired hardware architecture [J]. Journal of Electronic Measurement and Instrument, 2010, 24: 61 - 65. (in Chinese)

[5] 刘艳平, 沈慧芳, 韩凤霞. 医学细胞生物学[M]. 长沙: 中南大学出版社, 2001.

LIU Yanping, SHEN Yunfang, HAN Fengxia. Medical cell biology[M]. Changsha: Center South University Press, 2001. (in Chinese)

[6] Mange D, Sanchez E, Stauffer A. Embryonics: a new methodology for designing field-programmable gate arrays with self-repair and self-replicating properties [J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 1998, 6(3): 387 - 399.

[7] 周贵峰. 基于胚胎型细胞电路的 FIR 滤波器仿生自修复技术研究[D]. 长沙: 国防科学技术大学, 2010.

ZHOU Guifeng. Research on bio-inspired self-repairing technique of FIR filter via embryonic circuit[D]. Changsha: National University of Defense Technology, 2010. (in Chinese)

[8] Wolfram S. Theory and applications of cellular automata[C]// World Scientific, 1986.

[9] 周长林, 查永喜, 徐旭东, 等. 微生物学[M]. 北京: 中国医药科技出版社, 2003.

ZHOU Changlin, ZHA Yongxi, XU Xudong, et al. Microbiology [M]. Beijing: China Medical Science and Technology Press, 2003. (in Chinese)

[10] 田云, 徐文波, 张延伟. 无线通信 FPGA 设计[M]. 北京: 电子工业出版社, 2008.

TIAN Yun, XU Wenbo, ZHANG Yanwei. Wireless communication FPGA design[M]. Beijing: Publishing House of Electronics Industry, 2008. (in Chinese)

星载 SAR 天线阵面形变分析与补偿方法*

曾祥能^{1,2}, 何峰², 张永顺¹, 董臻²

(1. 空军工程大学 导弹学院, 陕西 三原 713800; 2. 国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘要: 研究了星载 SAR 天线阵面形变对波束输出的影响及形变补偿方法。针对未来星载 SAR 将采用的柔性阵面, 提出了一种用于星载 SAR 天线的空间形变实时测量与控制的闭环系统, 建立了阵面形变下阵列流形误差模型, 得出小幅度形变主要影响波束的旁瓣输出, 通过求解补偿形变权值的最小二乘解, 使阵列形变补偿后波束输出与期望波束输出最佳逼近, 并给出用于阵列误差补偿的阵列形变测量精度要求。仿真结果验证了本文方法结论的正确性与有效性。

关键词: 星载 SAR; 阵面形变; 误差补偿; 阵列流形

中图分类号: TN957.52 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0158-06

Analysis and compensation of spaceborne SAR antenna array deformation

ZENG Xiangneng^{1,2}, HE Feng², ZHANG Yongshun¹, DONG zhen²

(1. Missile Institute, Air Force Engineering University, Sanyuan 713800, China;

2. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The beam output of deformed Spaceborne SAR antenna array and the deformation error compensation are studied. Considering the flexible antenna array which may be adopted in the future Spaceborne SAR, a closed loop system for measuring and controlling of Spaceborne SAR antenna array's deformation is proposed, and the steering vector model of deformed antenna array is presented. It was found that the mild deformation affect the beam's sidelobe output more than the beam's main-lobe. Then, the compensated weights were solved by the least square method, which can approximate the deformed antenna array outputs to the original antenna array outputs at most. Next, the deformed antenna array measuring precise demand was worked out. The validity of this method is verified by simulation.

Key words: spaceborne SAR; antenna array deformation; deformation error compensation; steering vector

未来星载 SAR 将朝着轻量化、经济型、多功能、敏捷快响的方向发展^[1-2], 一般会使用可展开折叠的平面状天线, 其口径一般在 10~20m², 甚至在 30m² 以上。为了进一步实现轻量化设计、降低发射难度与成本, 星载 SAR 天线将采用轻质结构、并去掉刚性支架^[2]。因此, 星载 SAR 天线结构出现跨度大、质量轻、刚度低、阻尼弱等动力学特性, 属于大型挠性空间结构。在空间中, 由于卫星姿态调整、变轨、空间碎片碰撞、热载荷等干扰作用, 星载 SAR 天线会出现难以避免的空间位置形变, 造成天线波束畸变, 影响天线输出性能, 将会对星载 SAR 的模糊抑制性能产生严重影响, 造成阵列的信噪比损失, 对星载 SAR 天线阵面形变对波束输出的影响分析及误差补偿产生干扰。

针对 SAR 天线阵面形变对波束输出的影响, 学者们已经作了一定的研究, Wang 分析了天线的

典型机械误差对方向图函数的影响^[3]。陈杰等建立了天线展开误差模型及天线热变形误差模型, 基于成对回波理论分析了星载 SAR 相控阵天线展开误差和热变形误差对模糊性能的影响。Schipper 等以无人机载 SAR 天线为对象, 研究了天线变形或振动对其性能的影响, 并针对天线阵面的振动误差输出进行补偿^[4-6], 这些研究都是基于规律性阵列误差进行研究, 并重点研究天线阵列误差模型的构建。Takshshi 等分析了热振动对相控阵天线的影响, 并研究了在轨校准方法^[7]。Yonezawa 研究了展开相控阵天线的形变测量问题, 利用多个观测点对阵面形变进行实时观测并用于波束校正^[8]。加拿大的 MDA 公司对去刚性支架的柔性阵列天线进行了开创性的应用研究, 其通过一个专门的天线形变测量系统对柔性阵列天线的阵面形变进行实时测量, 再通过相

* 收稿日期: 2011-07-28
基金项目: 国家部委资助项目
作者简介: 曾祥能(1983—), 男, 湖南新宁人, 博士研究生, E-mail: zengxiangneng103@163.com;
张永顺(通信作者), 男, 教授, 博士生导师, E-mail: zhangyongshun@126.com

位补偿技术以保证阵列波束不变形^[9]。当前已有的研究主要集中在对天线振动、热变形造成的位置误差监测和影响分析,而对误差补偿的研究则主要是关于接收波束对散射目标的波达方向估计误差校正。对于星载 SAR 阵列天线,其发射波束的畸变将会影响测绘带内能量分布及模糊抑制性能,当前还很少有这方面的研究。本文基于小幅度的一维线阵天线形变问题,对星载 SAR 阵列天线的位置误差进行建模,分析其对波束输出产生的影响,并研究通过相位补偿权值来获得无误差波束输出。

1 阵列天线位置误差模型

当阵列天线的一个或多个阵元位置发生改变,相对于原有的阵列构型,由规则阵变为不规则阵,则阵列天线的方向图将发生改变。在一定程度上,位置误差和相位误差一样,原理上都是改变了原阵列的阵列流形,可按相位误差进行讨论。对于柔性阵列天线单元的位置误差测量,可通过在阵列单元的附着表明设置若干标志点,每个标志点处发射专用测量信号至平台上的位置测量系统,由多个无源接收系统对标志点进行空间位置解算,最后根据各天线单元的位置形变量进行相位补偿,使发射波束维持设定的形状,整个处理流程如图 1 所示。

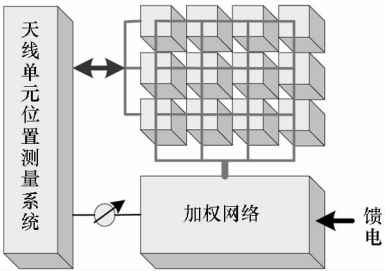


图1 阵列天线位置误差测量与补偿控制示意图

Fig.1 The configuration of antenna array cell position measuring and compensation

由 M 个阵元组成的均匀线阵,将阵列发射-接收角度空间表示为离散的角度集合 $\{\theta_n\}$ ($n = 1, \dots, N$), 天线阵列流形为 \mathbf{A} :

$$\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_n), \dots, \mathbf{a}(\theta_N)] \quad (1)$$

$\mathbf{a}(\theta_n)$ 为 $M \times 1$ 维指向 θ_n 的导向矢量:

$$\mathbf{a}(\theta_n) = \left[\exp\left(j2\pi \frac{d_1 \sin \theta_n}{\lambda}\right), \dots, \exp\left(j2\pi \frac{d_M \sin \theta_n}{\lambda}\right) \right]^T \quad (2)$$

式中 λ 为信号波长, d_m ($m = 1, \dots, M$) 为第 m 个阵元位置。

在阵列加权矢量 $\mathbf{W} = [w_1, \dots, w_M]^T$ 作用下

的发射波束输出为:

$$\mathbf{P} = \mathbf{W}^T \cdot \mathbf{A} \quad (3)$$

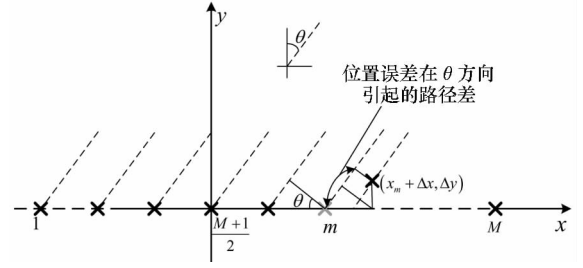


图2 线阵阵元位置误差引起路径差

Fig.2 The path difference arise from antenna cell position error of linear array

均匀线阵的二维形变如图2所示,第 m 个阵元位置由 $(x_m, 0)$ 变为 $(x_m + \Delta x_m, \Delta y_m)$ 后,在空间指向 θ 处路径差变化为

$$\Delta D_m(\theta) = \Delta x_m \cdot \sin \theta + \Delta y_m \cdot \cos \theta \quad (4)$$

显然,阵元位置相对原规则布阵发生改变后,其传播路径的改变量与空间指向 θ 相关,从而使阵列流形附加一个随 θ 变化的相位项。阵列位置误差改变了天线阵列流形,表示为 $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}(\theta_1), \dots, \tilde{\mathbf{a}}(\theta_n), \dots, \tilde{\mathbf{a}}(\theta_N)]$:

$$\tilde{\mathbf{a}}(\theta_n) = \tilde{\mathbf{T}}(\theta_n) \mathbf{a}(\theta_n) \quad (5)$$

式中 $\tilde{\mathbf{T}}(\theta_n)$ 为方位依赖的 $M \times M$ 维对角阵,其第 m 个对角元素 $[\tilde{\mathbf{T}}(\theta_n)]_m$ 对应第 m 个阵元在方位 θ_n 上的相位误差,可按下式计算:

$$\begin{aligned} [\tilde{\mathbf{T}}(\theta_n)]_m &= \exp\left[j2\pi \cdot \frac{\Delta D_m(\theta_n)}{\lambda}\right] \\ &= \exp\left[\frac{j2\pi}{\lambda} \cdot (\Delta x_m \cdot \sin \theta_n + \Delta y_m \cdot \cos \theta_n)\right] \end{aligned} \quad (6)$$

则阵列发生形变后,其指向为 θ_0 的均匀权发射波束输出为

$$\tilde{\mathbf{P}} = \mathbf{W}^T \cdot \tilde{\mathbf{A}} \quad (7)$$

阵列流形的改变量为 $\Delta \mathbf{A} = \tilde{\mathbf{A}} - \mathbf{A}$, 对应的输出发射波束形状改变量为

$$\begin{aligned} \Delta \mathbf{P} &= \tilde{\mathbf{P}} - \mathbf{P} \\ &= \mathbf{W}^T \cdot \Delta \mathbf{A} \\ &= \mathbf{W}^T \cdot \{[\tilde{\mathbf{T}}(\theta_1) - \mathbf{I}] \cdot \mathbf{a}(\theta_1), \dots, \\ &\quad [\tilde{\mathbf{T}}(\theta_N) - \mathbf{I}] \cdot \mathbf{a}(\theta_N)\} \end{aligned} \quad (8)$$

则阵列形变引起的 θ_n 方向的波束输出改变量 $\Delta \mathbf{P}(\theta_n)$ 为

$$\begin{aligned} \Delta \mathbf{P}(\theta_n) &= \mathbf{W}^T \cdot [\tilde{\mathbf{T}}(\theta_n) - \mathbf{I}] \cdot \mathbf{a}(\theta_n) \\ &= \sum_{m=1}^M [\mathbf{W}^T]_m \cdot ([\tilde{\mathbf{T}}(\theta_n)]_m - 1) \cdot [\mathbf{a}(\theta_n)]_m \end{aligned} \quad (9)$$

由于在发射波束的主瓣区域内满足 $[\mathbf{W}^T]_m \cdot$

$[\mathbf{a}(\theta_n)]_m \approx 1, (m = 1, \dots, M)$ 。则对于主瓣区域 Θ_{main} 内的波束输出改变量可近似表示为

$$\Delta \mathbf{P}(\theta) = \text{tr}(\tilde{\mathbf{I}}(\theta)) - M, \quad (\theta \in \Theta_{\text{main}}) \quad (10)$$

其中 $\text{tr}(\cdot)$ 表示取矩阵的迹, 由式(6)、(10)可知, 当形变的阵元数较少且形变量 ΔD_m 较小时, 有 $\sum_{m=1}^M [\tilde{\mathbf{I}}(\theta)]_m \approx M$, 即发射波束的主瓣区域波束输出改变量较小。在旁瓣区域则不同, 由式(9)可见, 在旁瓣区域可能出现 $[\mathbf{W}^T]_m \cdot ([\tilde{\mathbf{I}}(\theta_n)]_m - 1) \cdot [\mathbf{a}(\theta_n)]_m = 1$, 从而使波束输出的改变剧烈。因此, 柔性天线的小范围微量形变主要影响发射波束的旁瓣输出, 对主瓣输出的改变量并不显著。

2 阵列天线位置误差补偿

为了维持无阵列形变时的阵列波束形状, 则考虑调整阵列天线发射波束对应的权值, 使阵列波束输出尽量接近无形变的波束输出。存在形变的阵列流形为 $\tilde{\mathbf{A}}$, 设用于补偿阵列天线位置误差的阵列权值为 \mathbf{W}_{com} , 则经此权值补偿后的波束输出为

$$\begin{aligned} \mathbf{P}_{\text{com}} &= \mathbf{W}_{\text{com}}^T \cdot \tilde{\mathbf{A}} \\ &= \mathbf{W}_{\text{com}}^T \cdot [\tilde{\mathbf{I}}(\theta_1)\mathbf{a}(\theta_1), \dots, \tilde{\mathbf{I}}(\theta_N)\mathbf{a}(\theta_N)] \end{aligned} \quad (11)$$

将式(4)与式(3)对比, 则要使完全补偿掉阵列天线位置误差, 就要满足 $\mathbf{P}_{\text{com}} = \mathbf{P}$, 即:

$$\mathbf{W}_{\text{com}}^T \cdot \tilde{\mathbf{I}}(\theta_n)\mathbf{a}(\theta_n) = \mathbf{P}(\theta_n) \quad (n = 1, \dots, N) \quad (12)$$

一般的, $N > M$, 则式(12)为超定方程组, 无法求得整个方程组的解析解, 但可求得 \mathbf{W}_{com} 的最小二乘解:

$$\mathbf{W}_{\text{com}} = (\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\tilde{\mathbf{A}}\mathbf{P}^T \quad (13)$$

若阵列波束的离散角度数量较大, 则可对式(12)中的离散角度数 N 中取重点区域的角度区域(如主瓣区域)进行求解。在此基础上, 若局部旁瓣水平过高, 可通过对已获得的权值 \mathbf{W}_{com} 进行微扰, 形成新的阵列控制权值 $\mathbf{W}_{\text{com}} + \Delta \mathbf{W}$, 实现对局部旁瓣区域的控制:

$$\begin{aligned} \min \quad & \Delta \mathbf{W}^H \Delta \mathbf{W} \\ \text{s. t.} \quad & |(\mathbf{W}_0 + \Delta \mathbf{W})^H \mathbf{a}(\theta_s)| \leq \varepsilon \\ & |\Delta \mathbf{W}^H \mathbf{a}(\theta_m)| \leq \delta \end{aligned} \quad (14)$$

其中 θ_s 表示旁瓣区域, θ_m 表示主瓣区域, ε 的设定根据所期望的旁瓣水平而定, δ 为可容忍的主瓣能量波动范围。

当阵列形变量比较小时, $\tilde{\mathbf{A}}$ 相对于 \mathbf{A} 的结构改变比较小, 如每个单元的位置形变量较小, 或者只有少数几个单元的位置形变量较大, 这时, 最小二乘解 $\mathbf{W}_{\text{com}}^T$ 获得波束输出与无形变的波束输出具有最小均方误差意义, 二者很接近。但是, 当阵列天线形变剧烈时, $\tilde{\mathbf{A}}$ 相对于 \mathbf{A} 的结构改变严重, 则通过求取 \mathbf{W}_{com} 的最小二乘解已不能满足阵列位置误差补偿的要求, 波束输出形状畸变严重。对于柔性阵列天线的形变严重问题, 将通过机械控制调整阵列天线单元的空间布阵来解决。

3 阵面形变测量精度要求

根据式(13)得阵列形变补偿校正后波束与无阵列形变的波束输出具有最小二乘意义上的一致, 但相位补偿量计算是基于对阵列形变测量值获得的, 因此, 形变测量精度仍将影响着补偿后的波束输出。测量精度是指测量误差分布的离散与密集程度、与观测条件有关。评定测量精度的常用指标包括中误差、相对误差和极限误差(容许误差), 中误差与统计学的标准差定义形式相同:

$$\sigma = \pm \sqrt{\frac{\Delta_1^2 + \Delta_2^2 + \dots + \Delta_n^2}{n}} \quad (15)$$

式中 Δ_i 为第 i 个形变量测量值的偶然误差。

记阵列形变的测量精度为 σ , 对阵列形变量的测量值 $\Delta \tilde{\mathbf{D}}$ 为

$$\Delta \tilde{\mathbf{D}} = \Delta \mathbf{D} + \sigma \quad (16)$$

以 M 个单元的线阵天线为例, 阵列形变补偿后的波束输出受形变测量精度的影响, 可理解为经形变补偿后的天线阵列仍存在幅度为 $\sigma = [\sigma_1, \dots, \sigma_M]$ 的位置形变, σ 为各元素独立的 M 维向量, 由于对天线阵列形变的测量分别在 x 坐标方向和 y 坐标方向进行, 且对 M 个阵列单元的形变测量条件相同, 则测量值中误差可分为 σ_x 、 σ_y 两类, 根据误差传播定律的线性函数误差传播定律, 容易得到 $\Delta D_m(\theta) = \Delta x_m \cdot \sin \theta + \Delta y_m \cdot \cos \theta$ 的中误差 $m_{\Delta D}$ 为

$$m_{\Delta D} = \sqrt{\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta} \quad (17)$$

进一步的, 可得 $[\mathbf{I}(\theta)]_m = \exp\left(\frac{j2\pi}{\lambda} \Delta D_m(\theta)\right)$ 的中误差 m_{Γ_m} 为

$$\begin{aligned} m_{\Gamma_m} &= \frac{2\pi}{\lambda} [\mathbf{I}(\theta)]_m \cdot m_{\Delta D} \\ &= \frac{2\pi}{\lambda} [\mathbf{I}(\theta)]_m \cdot \sqrt{\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta} \end{aligned} \quad (18)$$

根据式(11)可得形变补偿后的波束输出为

$$P_{\text{com}}(\theta) = \sum_{m=1}^M [\mathbf{W}_{\text{com}}^T]_m \cdot [\mathbf{a}(\theta)]_m \cdot [\mathbf{\Gamma}(\theta)]_m \quad (19)$$

$$\begin{aligned} m_{P_{\text{com}}}(\theta) &= \sqrt{\sum_{m=1}^M ([\mathbf{W}_{\text{com}}^T]_m \cdot ([\mathbf{a}(\theta)]_m)^2 \cdot m_{\Gamma_m}^2)} \\ &= \sqrt{\sum_{m=1}^M ([\mathbf{W}_{\text{com}}^T]_m \cdot [\mathbf{a}(\theta)]_m)^2 \cdot [\mathbf{\Gamma}(\theta)]_m^2 \cdot \frac{4\pi^2}{\lambda^2} \cdot (\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta)} \quad (20) \\ &= \frac{2\pi P_{\text{com}}(\theta)}{\lambda} \cdot \sqrt{\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta} \end{aligned}$$

则阵列形变测量精度 (σ_x, σ_y) 下波束输出中误差为 $\frac{2\pi P_{\text{com}}(\theta)}{\lambda} \cdot \sqrt{\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta}$,其表示在阵列形变测量精度 (σ_x, σ_y) 的情况下,在角度 θ 方向出现的极限误差,其值与波束输出的空间角度方向有关。进一步的,将形变测量精度引起的波束输出容许误差折算成增益损失相对值 $\eta =$

$$\frac{m_{P_{\text{com}}}}{P_{\text{main}}}(P_{\text{main}} \text{ 为波束主瓣电平})。在主瓣区域,有$$

$$\eta \approx \pm \frac{2\pi}{\lambda} \cdot \sqrt{\sigma_x^2 \cdot \sin^2 \theta + \sigma_y^2 \cdot \cos^2 \theta} \quad (21)$$

通常对 (x, y) 坐标方向的阵列形变测量条件是相同的,可以取测量精度 (σ_x, σ_y) 为统一的精度 σ 。主瓣增益损失随测量精度误差的变化趋势如图3所示。可见,当测量精度误差 $\sigma \geq 0.1592\lambda$ (λ 为电磁波长)时,波束输出的误差相对值为1,此时由测量精度引起的波束输出误差可能导致波束完全变形失效。如要求形变测量误差引起波束主瓣增益损失低于1dB,可以计算出 $\sigma \leq 0.0173\lambda$ 。

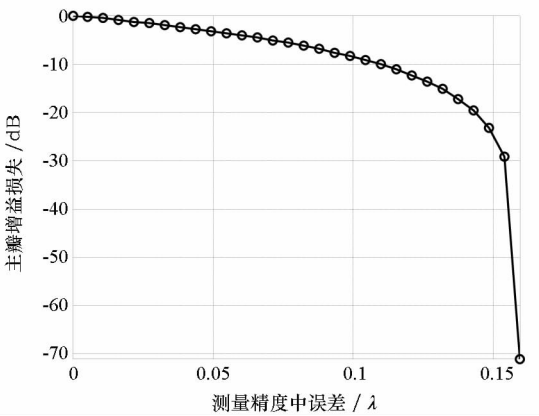


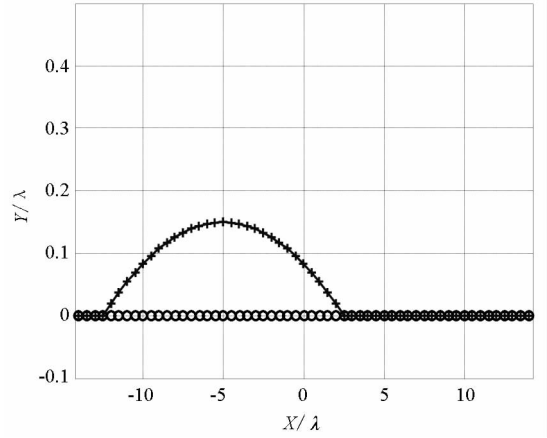
图3 主瓣增益损失与测量精度误差的关系
Fig.3 The relationship between main-lobe loss and measuring precision error

4 仿真分析

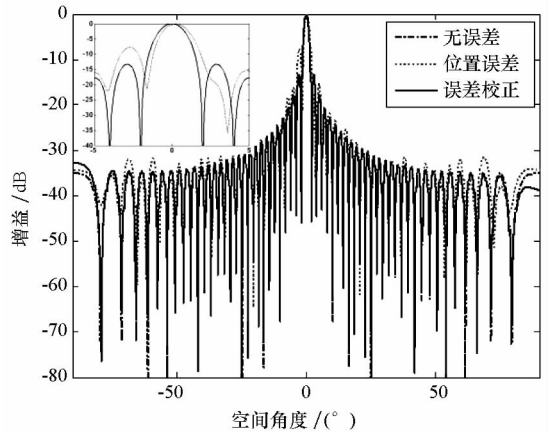
(1)由57个相同全向阵元组成的均匀线阵,

则可得 $P_{\text{com}}(\theta)$ 的中误差 $m_{P_{\text{com}}}$ 如式(20)所示。

如图4(a)所示,阵列的单元间距为 0.5λ ,在第5~34个单元出现最大幅度为 0.15λ 的扭曲凸起变形,其波束输出情况如图4(b)所示,可见,阵列形变的位置误差主要引起波束输出的旁瓣情况,主瓣改变较小。通过采用本文方法进行位置误差补偿校正后,其波束输出与无误差情况基本一致。



(a)阵列形变前后情况



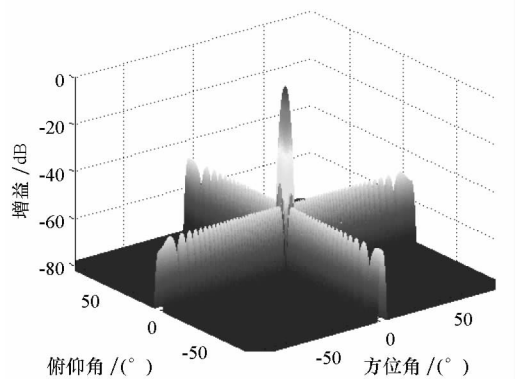
(b)阵列形变补偿前后波束输出情况

图4 线阵天线形变补偿情况

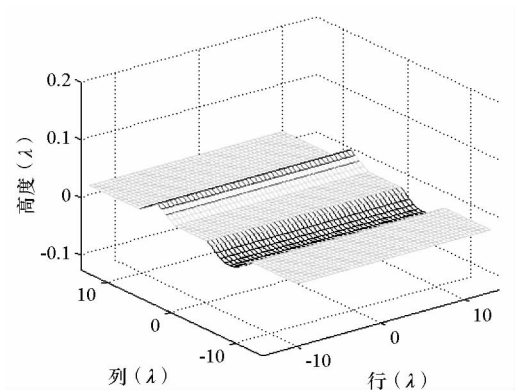
Fig.4 The distortion compensation of uniform linear antenna array

(2)由 47×47 个各向同性单元组成的均匀面阵,阵列行间距为 0.5λ ,列间距为 0.7λ ,采用二维切比雪夫幅度权控制旁瓣水平为 -50dB ,波束

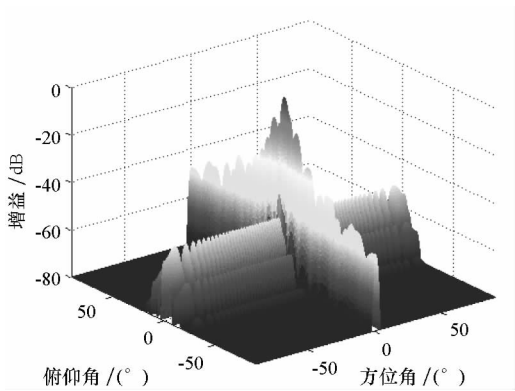
输出如图 5(a) 所示。沿列方向出现幅度分别为 0.1λ 、 0.05λ 的二维褶皱形变,如图 5(b) 所示,阵面形变导致天线波束畸变如图 5(c) 所示,采用本文方法对阵面沿列方向二维位置误差进行相位补偿校正,校正后波束输出如图 5(d) 所示,可见,校正后波束输出与误差时一致。



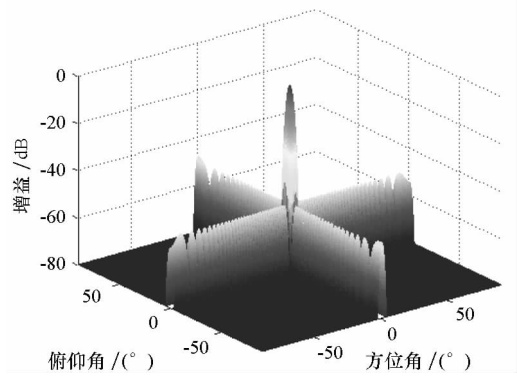
(a) 均匀阵面方向图



(b) 阵面二维形变



(c) 阵面形变引起波束畸变



(d) 误差补偿校正后波束输出

图 5 均匀面阵形变补偿情况

Fig.5 The distortion compensation of uniform planar antenna array

(3) 现考察本文方法在不同形变程度下,经误差补偿后的波束输出情况,阵列形变幅度为 $0 \sim 1\lambda$,由于阵列波束输出情况主要影响天线的模糊抑制性能,因此波束输出情况衡量指标选为综合旁瓣水平,其定义为旁瓣增益与主瓣增益之比。通过 100 次蒙特卡罗试验,得误差补偿校正后输出综合旁瓣水平与形变幅度关系如图 6 所示,经过本文方法作误差补偿后,波束主瓣情况与无误差情况基本接近,但部分旁瓣增益抬高严重,随着形变幅度增加,误差补偿后波束旁瓣畸变剧烈,当形变幅度达 0.5λ 时,综合旁瓣增益水平被抬高至 -8dB 以上,因此,本文方法在形变幅度为 0.5λ 内有效。

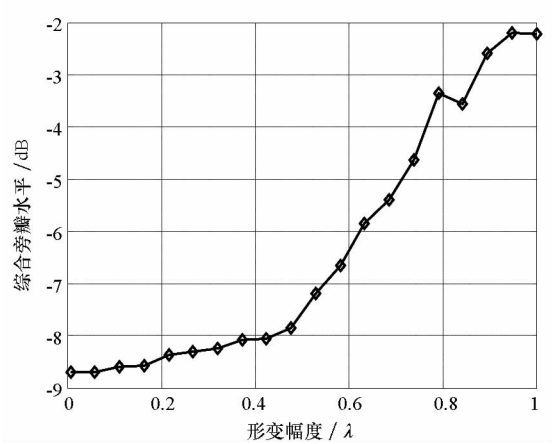


图 6 误差补偿校正后输出综合旁瓣水平与形变幅度的关系

Fig.6 The relationship between ISRL and distortion amplitude of distortion compensation beam

5 结 论

文章对一维线阵的阵列形变问题进行了研究,建立了阵列天线位置误差模型,分析出较小幅度的天线形变主要影响波束旁瓣输出,对主瓣输

出的影响较小。然后,对阵列天线的位置误差通过相位幅权值进行补偿,获得了无误差的波束输出性能。本文方法面向未来星载 SAR 的柔性阵列天线的实时波束控制,对阵列天线形变的实时测量将通过专门的形变测量系统实时获得,研究的结论适用于一维线阵的阵列形变问题,也适用于二维面阵的仅一维形变时的误差补偿。当阵列形变剧烈时,可通过机械控制、膨胀塑型等手段调整阵列的空间位置,使其复原到接近无误差阵列,这时就又回到了本文讨论的小幅度天线形变问题。

参考文献 (References)

[1] Keydel W. Perspectives and visions for future SAR systems[J]. IEE Pvoc. -Radar Sonar Nuvig. , 2003, 150(3): 97 – 103.

[2] James K, Fox P, Thompson A. WiSAR™: A disruptive technology for high-performance, smallsat-based synthetic aperture radar missions [C]// 6th Responsive Space Conference. 2008. Los Angeles, CA. 1 – 7.

[3] Wang H S C. Performance of phased-array antennas with mechanical errors[J]. IEEE Transactions on Aerospace and Electronic Systems, 1992, 28(2):535 – 545.

[4] van Tongeren J H, Schippers H, Vos G. Development of smart antennas on vibrating structures of aerospace platforms[C]// NATO AVT Specialists Meeting, Vilnius, Lithuania, 2006: 1 – 16.

[5] van Tongeren J H, Schippers H, Knott P, et al. Vibrating antennas and compensation techniques Research in NATO/ RTO/SET 087/RTG 50 [C]// IEEE Aerospace Conference. 2007: 1 – 13.

[6] Vos G, Schippers H. Analysis of vibrating lightweight radar antennas [C]// International Radar Conference, 2009. Bordeaux. 1 – 4.

[7] Nakamoto N, Takahshi T, Ohtsuka M, et al. A simple on-board calibration method and its accuracy for mechanical distortions of satellite phased array antennas [C]. EuCAP 2009: 1573 – 1577.

[8] Konishi Y, Yonezawa K, Chiba I, Beam-shape correction in deployable phased arrays[J]. IEEE transactions on antennas and propagation, 1999, 47(3):482 – 486.

[9] James K, Peterman D, Glavac V. Distortion measurement and compensation in a synthetic aperture radar phased-array antenna [C]// 2010 14th International Symposium on Antenna Technology and Applied Electromagnetics and the American Electromagnetics Conference, 2010. American IEEE: 978 – 982.

星载干涉仪无源定位新方法及其误差分析*

李 腾,郭福成,姜文利

(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘 要: 针对传统单星二维干涉仪测向定位存在的设备量大、易受通道间幅度/相位不一致性影响等缺点,提出了在单颗自旋卫星上只安装两个接收通道构成一维干涉仪,测量相位差变化率的无源定位新方法,采用了基于粒子群优化(PSO)的定位算法,分析了干涉仪转速以及旋转平面对定位性能的影响。结果表明,增大干涉仪转速有利于提高定位精度,为了使星下点周围各个方向上都有较好的定位精度,干涉仪旋转平面应与初始观测时刻卫星位置矢量垂直,PSO 算法的定位精度能够接近定位误差的克拉美罗下限(CRLB)。

关键词: 无源定位;干涉仪;自旋卫星;相位差变化率

中图分类号: TN97 **文献标志码:** A **文章编号:** 1001-2486(2012)03-0164-07

A novel method for satellite-borne passive localization
using interferometer and its error analysis

LI Teng, GUO Fucheng, JIANG Wenli

(College of Electronic Science and Engineering, National University
of Defense Technology, Changsha 410073, China)

Abstract: The conventional line-of-sight (LOS) geo-location method using two-dimension interferometer on a satellite leads to heavy payloads and it is always under the effect of amplitude-phase inconsistencies among receiving channels. A novel passive localization method is proposed, which merely needs two receiving channels to form a one-dimension interferometer on a spin-stabilized satellite. Rate of phase difference change was measured for emitter localization. A location algorithm based on particle swarm optimization (PSO) was designed. The impact of rotating speed and rotating plane of the interferometer on localization precision was analyzed. The results showed that localization precision can be improved by increasing rotating speed of the interferometer, the rotating plane should be perpendicular to the satellite position vector at the initial moment in order to get good localization precision in all directions around satellite projection point on the ground, and localization precision of the proposed algorithm can approach the Cramer-Rao lower bound (CRLB).

Key words: passive localization; interferometer; spin-stabilized satellite; rate of phase difference change

利用单颗卫星携带的电子设备对地面无线电辐射源进行定位具有隐蔽性好、覆盖范围广、数据同步简单等优点^[1],是一个很有意义的研究方向。理论上,利用单个接收通道即可实现单星对目标的定位,如徐义等人提出的单星仅测脉冲到达时间(TOA)无源定位法^[1]、陆安南等人提出的单星仅测频无源定位法^[2]以及张敏等人提出的单星测多普勒变化率无源定位法^[3],这些方法的本质都是利用相对运动引起的多普勒信息定位,虽然所需设备量少,但是对信号形式的要求比较严格,如分别要求发射信号的脉冲重复频率固定^[1]和载频固定^[2-3]。而基于干涉仪体制的无源定位系统,可以利用通道间接收信号的互相关消除信号波形的影响,对信号形式的适应能力强,且可以利用长基线提高性能增益^[4],因此更具有工

程实用价值。

现有基于干涉仪的单星无源定位方法中,常用的是二维干涉仪测向定位方法,即利用二维干涉仪测得的方向线与地面相交进行定位^[5-6]。该方法能够实现单次测量定位,但存在两个主要问题:一是所需设备大,至少需要3个接收通道,且增大基线长度时,通常还需要增加接收通道来解相位差模糊^[7-8];二是通道间的幅度/相位不一致性会给角度测量和定位带来较大误差^[9],且通道数越多,该不一致性越难校正。为了克服干涉仪测向定位的这两个缺点,郭福成等人在机载无源定位应用背景下提出了利用两通道干涉仪只测相位差变化率的无源定位方法,研究结果表明通过载机机动带动干涉仪姿态变化能够带来定位精度的大幅提升^[10],但是他们没有对这一结果进行深

* 收稿日期:2011-07-06

作者简介:李腾(1984—),男,河南信阳人,博士研究生,E-mail:liteng@nudt.edu.cn;

姜文利(通信作者),男,教授,博士,博士生导师,E-mail:jiangwl@nudt.edu.cn

入的理论分析,也没有给出实用的定位算法。

受文献[10]的启发,考虑到自旋卫星的高速旋转特性,本文提出一种在自旋卫星上仅安装两个接收通道构成一维干涉仪,测量接收信号相位差变化率对目标定位的新方法,给出了定位求解算法,并深入分析了干涉仪转速以及旋转平面对定位性能的影响。相位差变化率为两个通道间相位差的一阶导数,因此利用相位差变化率定位可以在很大程度上消除通道间固定偏差的影响,且当接收信号连续时,估计相位差变化率一般不需要解相位差的绝对模糊^[11],因此可以在不增加设备量的情况下增大基线长度以提高性能增益。

1 数学建模与定位求解

1.1 数学模型

如图1所示, T 为地球表面的辐射源,发射波长为 λ 的电磁波, S_1 和 S_2 为安装在自旋卫星上的两个接收天线,二者的连线垂直于卫星自旋轴,构成基线长度为 l 的干涉仪,基线波长比因子 $K=2\pi l/\lambda$, S 为星下点, xyz 为地心惯性直角坐标, T 的坐标为 x_T 。

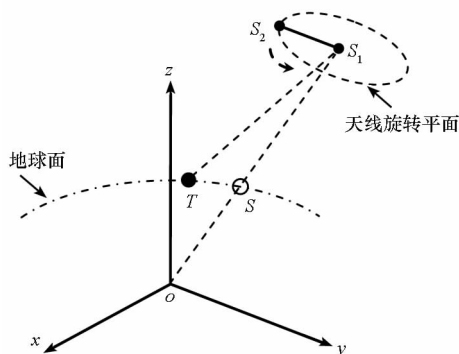


图1 定位场景示意图

Fig. 1 Sketch map of the localization scenario

在其他因素相同时,定位精度取决于目标和卫星的相对位置关系而非二者的绝对位置,所以在分析定位误差分布时应该采用目标与星下点轨迹的相对位置关系来描述,但是该位置关系没有现成的物理量来表述,因此为了避开复杂的坐标变换,本文假设初始观测时刻卫星的经、纬度均为 0° ,卫星轨道为赤道圆轨道,这样经度变化方向即代表沿星下点轨迹的方向,纬度变化方向即代表垂直于星下点轨迹的方向。另外,为了简化公式表述,取正球面地球模型。

定义干涉仪坐标系 $x'y'z'$,其坐标原点为 S_1 在初始观测时刻的位置, $x'y'$ 平面为干涉仪旋转平面, x' 轴正向为初始时刻干涉仪基线方向, y' 轴正向为在 $x'y'$ 平面内垂直于 x' 轴的方向,按照右

手定则确定 z' 轴方向。 H 为 $x'y'z'$ 系到 xyz 系的矢量旋转矩阵。

自旋卫星的姿态控制一般间隔进行,不妨假设在一次定位过程中不对卫星进行姿态控制,因此可认为卫星自旋轴在空间的指向保持不变^[12],即干涉仪旋转平面始终平行于 $x'y'$ 平面。

在 m 时刻, b_m 和 b'_m 分别表示 xyz 系和 $x'y'z'$ 系下干涉仪基线单位矢量, α'_m 表示 b'_m 在 $x'y'z'$ 系的方位角, ω_m 表示干涉仪旋转角速度, x_{0m} 和 \dot{x}_{0m} 分别表示 xyz 系下卫星的位置矢量和速度矢量,则

$$b_m = Hb'_m \quad (1)$$

$$b'_m = [\cos\alpha'_m, \sin\alpha'_m, 0]^T \quad (2)$$

令 $x_m = x_T - x_{0m}$, $\dot{x}_m = -\dot{x}_{0m}$,则目标到卫星的距离为

$$r_m = \|x_m\| = \sqrt{x_m^T x_m} \quad (3)$$

式中 $\|\cdot\|$ 表示矢量的长度,干涉仪两天线接收信号的相位差为

$$\varphi_m = -K\cos\theta_m \quad (4)$$

式中 θ_m 为干涉仪基线与目标视向的夹角,

$$\cos\theta_m = x_m^T b_m / r_m \quad (5)$$

将(5)式代入(4)式,并对(4)式求导得相位差变化率为

$$f_m(x_T) = K \left(\frac{(x_m^T \dot{x}_m)(x_m^T b_m)}{r_m^3} - \frac{\dot{x}_m^T b_m + x_m^T (\omega_m H A b'_m)}{r_m} \right) \quad (6)$$

其中

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

实际测量的相位差变化率为

$$\dot{\varphi}_m = f_m(x_T) + \varepsilon_m \quad (8)$$

式中 ε_m 为测量误差。

假设在一次定位的过程中共进行了 M 次相位差变化率测量,构成观测矢量

$$\dot{\varphi} = [\dot{\varphi}_1, \dots, \dot{\varphi}_M]^T \quad (9)$$

对于任意一点 x ,令

$$f(x) = [f_1(x), \dots, f_M(x)]^T \quad (10)$$

根据最大似然准则可得目标位置的估计值。假设各次观测误差的方差相同,最大似然估计等价于求解使得目标函数

$$C = \|\dot{\varphi} - f(x)\|^2 \quad (11)$$

达到最小的 x ,用公式可表示为

$$\hat{x}_T = \min_x \|\dot{\varphi} - f(x)\|^2 \quad (12)$$

1.2 基于粒子群优化(PSO)的定位算法

对(12)式的求解是一个多峰函数寻优问题,

常用的寻优算法,如拟牛顿法、梯度搜索法等对初值比较敏感^[13],网格搜索法虽然没有初值设置问题,但是计算量很大^[14]。PSO 方法作为一种新的群智能方法,具有收敛速度快,通用性强等优点,且计算量不大^[15-18],因此本文采用标准 PSO 方法^[16]作为定位算法。该算法能够兼顾粒子的广度搜索与深度搜索能力^[18],且算法流程简单。

在 PSO 中,每个粒子 $\chi_i = [\chi_{i1}, \dots, \chi_{id}]$ 表示 D 维待估状态空间的一个点,下标 i 为粒子的序号,粒子速度为 $\dot{\chi}_i = [\dot{\chi}_{i1}, \dots, \dot{\chi}_{id}]$,粒子目前经过的最好位置用 $\mathbf{q}_i = [q_{i1}, \dots, q_{id}]$ 表示,所有粒子目前经过的最好位置用 $\mathbf{q}_g = [q_{g1}, \dots, q_{gd}]$ 表示。粒子速度和位置更新公式如(14)式所示。

$$\begin{cases} \chi_{id}^{k+1} = w\dot{\chi}_{id}^k + c_1 r_{1d}^k (q_{id}^k - \chi_{id}^k) + c_2 r_{2d}^k (q_{gd}^k - \chi_{id}^k) \\ \dot{\chi}_{id}^{k+1} = \dot{\chi}_{id}^k + \dot{\chi}_{id}^{k+1} \end{cases} \quad (13)$$

式中,下标 d 表示第 d 个待估状态分量,上标 k 为迭代步数, w 为惯性权重, c_1 和 c_2 为学习率, r_{1d} 和 r_{2d} 为 $[0, 1]$ 区间上的随机数。

下面给出本文定位算法的流程和主要参数。

Step 1 选取地球表面某一感兴趣的观测区域,一般为天线波束有效覆盖区域,本文选取初始时刻星下点经纬度各正负 10° 的区域为观测区域。

Step 2 设置最大迭代步数 400;初始惯性权重 0.9,前 100 次迭代时线性递减为 0.6,然后保持不变;粒子数目 30, c_1 和 c_2 均为 2.1。

Step 3 在观测区域内按照均匀分布产生初始的各个粒子 χ_i 及其速度 $\dot{\chi}_i$,并根据(6)~(12)式分别计算其代价函数,存入 C_i 。

Step 4 对于各个粒子,根据(6)~(12)式分别计算其代价函数。

Step 5 比较粒子的代价函数和 C_i ,如果当前代价函数更小,用它取代 C_i ,并将 χ_i 修正为当前位置。

Step 6 求得所有粒子中代价函数最小的粒子,将其下标存入变量 g 。

Step 7 根据(13)式更新粒子速度及位置。

Step 8 重复 step 4~step 7 直到满足停止准则,停止准则通常设置为满意的代价函数值或最大迭代次数。

Step 9 选取代价函数最小的粒子的位置作为目标位置的估计。

2 定位误差的 CRLB

目标位置满足地球面约束方程,因此对目标

位置的估计实际上是一个两参数估计问题,可将 \mathbf{x}_T 用目标的经度 L 和纬度 B 表示为

$$\mathbf{x}_T = R[\cos B \cos L, \cos B \sin L, \sin B]^T \quad (14)$$

式中 R 为地球半径。

根据(8)式,假设测量误差 $\varepsilon_1, \dots, \varepsilon_N$ 为独立同分布的零均值高斯白噪声,方差为 σ_φ^2 。令 $y = [L, B]^T$,则 $\dot{\varphi}$ 关于 y 的 Fisher 信息矩阵为

$$J(y) = E \left[\left(\frac{\partial \ln p(\dot{\varphi} | \mathbf{x}_T)}{\partial y} \right)^T \left(\frac{\partial \ln p(\dot{\varphi} | \mathbf{x}_T)}{\partial y} \right) \right] \quad (15)$$

$p(\dot{\varphi} | \mathbf{x}_T)$ 为 $M \times 1$ 的列矢量,表示 M 次观测的概率密度,其第 m 个元素为

$$p(\dot{\varphi}_m | \mathbf{x}_T) = \frac{1}{\sqrt{2\pi}\sigma_\varphi} \exp \left\{ -\frac{[\dot{\varphi}_m - f_m(\mathbf{x}_T)]^2}{2\sigma_\varphi^2} \right\} \quad (16)$$

将(14)、(16)式代入(15)式,整理可得

$$J(y) = \frac{1}{\sigma_\varphi^2} \sum_{m=1}^M \begin{bmatrix} (\mathbf{h}_m^T \mathbf{a}_L)^2 & (\mathbf{h}_m^T \mathbf{a}_L)(\mathbf{h}_m^T \mathbf{a}_B) \\ (\mathbf{h}_m^T \mathbf{a}_L)(\mathbf{h}_m^T \mathbf{a}_B) & (\mathbf{h}_m^T \mathbf{a}_B)^2 \end{bmatrix} \quad (17)$$

其中

$$\mathbf{a}_L = R[-\cos B \sin L, \cos B \cos L, 0]^T \quad (18)$$

$$\mathbf{a}_B = R[-\sin B \cos L, -\sin B \sin L, \cos B]^T \quad (19)$$

$$\begin{aligned} \mathbf{h}_m = & -\frac{K}{r_m} \{ \omega_m (\mathbf{b}_{m\perp} - \mathbf{u}_m^T \mathbf{b}_{m\perp} \mathbf{u}_m) - (\mathbf{u}_m^T \mathbf{b}_m) \rho_m \\ & - (\mathbf{u}_m^T \rho_m) \mathbf{b}_m - (\rho_m^T \mathbf{b}_m) \mathbf{u}_m + 3(\mathbf{u}_m^T \rho_m)(\mathbf{u}_m^T \mathbf{b}_m) \mathbf{u}_m \} \end{aligned} \quad (20)$$

其中

$$\mathbf{u}_m = \mathbf{x}_m / r_m, \rho_m = \dot{\mathbf{x}}_m / r_m, \mathbf{b}_{m\perp} = \mathbf{H} \mathbf{A} \mathbf{b}'_m \quad (21)$$

CRLB 为 Fisher 信息矩阵的逆矩阵,即

$$\text{CRLB}(\mathbf{y}) = \mathbf{J}^{-1}(\mathbf{y}) \quad (22)$$

则

$$\text{CRLB}(\mathbf{x}_T) = \mathbf{M} \cdot \text{CRLB}(\mathbf{y}) \cdot \mathbf{M}^T \quad (23)$$

其中

$$\mathbf{M} = [\mathbf{a}_L, \mathbf{a}_B] \quad (24)$$

对于不同的 \mathbf{x}_T ,根据(23)式可计算其相应的 CRLB,从而得定位误差的几何稀释(GDOP)为^[19]

$$\text{GDOP}(\mathbf{x}_T) = \sqrt{\text{trace}(\text{CRLB}(\mathbf{x}_T))} \quad (25)$$

式中 $\text{trace}(\cdot)$ 表示求矩阵的迹。

3 定位性能分析

CRLB 的表达式比较复杂且难以化简,因此通过 CRLB 的表达式直接进行理论分析比较困难。Kalaba 与 Spingarn 指出,通过提高非线性系统观测数据对于待估参数的敏感度,可以改善参数估计精度^[20],而该敏感度可以通过观测量对于待

估参数的导数来表征。对于本文所讨论的问题,定义敏感度因子

$$\eta_{mL} = |\mathbf{h}_m^T \mathbf{a}_L|, \quad \eta_{mB} = |\mathbf{h}_m^T \mathbf{a}_B| \quad (26)$$

由(17)式可知,在测量误差一定的情况下, $\dot{\varphi}$ 关于 y 的Fisher信息矩阵 $\mathbf{J}(y)$ 即取决于 η_{mL} 和 η_{mB} 。定性而言, $\eta_{mL}(\eta_{mB})$ 越大,观测方程对于 $L(B)$ 的约束越强,则在相同的相位差变化率误差下, $L(B)$ 的估计越准确。由于目标位置的估计由 L 和 B 的估计结果共同决定,所以只有当 L 和 B 都估计准确时,才能得到准确的定位结果。

卫星的运动速度与其到目标的距离的量值一般差距较大,并且轨道越高,差距越大,如对于600km轨道高度的卫星,其速度约为7.6km/s,此时 $\|\rho_m\| < 0.013$,而一般自旋卫星的转速较快,可保证 $\omega_m \gg \|\rho_m\|$,则(20)式中起主要作用的仅为不含 ρ_m 的项,因此,可近似得到

$$\mathbf{h}_m = -\omega_m K \mathbf{v}_m / r_m \quad (27)$$

其中

$$\mathbf{v}_m = \mathbf{b}_{m\perp} - \mathbf{u}_m^T \mathbf{b}_{m\perp} \mathbf{u}_m \quad (28)$$

当 $\mathbf{b}_{m\perp}$ 与 \mathbf{u}_m 垂直时, $\|\mathbf{v}_m\|$ 达到最大值1。由于在一次定位的短时间内,卫星到目标的距离 r_m 一般变化不大,可近似都用 r 表示,另外,考虑干涉仪随卫星自旋以角速度 ω 匀速转动的情况,将(27)式代入(26)式可得

$$\eta_{mL} = \omega K |\mathbf{v}_m^T \mathbf{a}_L| / r \leq \omega K \|\mathbf{a}_L\| / r \quad (29)$$

$$\eta_{mB} = \omega K |\mathbf{v}_m^T \mathbf{a}_B| / r \leq \omega K \|\mathbf{a}_B\| / r \quad (30)$$

由(29)和(30)式可知,在其他条件相同时, ω 越大, η_{mL} 和 η_{mB} 均越大,相同的相位差变化率误差引起的 L 和 B 的估计误差越小,从而定位精度越高。

根据卫星自旋轴指向的不同,干涉仪旋转平面存在各种情况,本文分析3种典型的情况,其他情况可以进行类似分析,3种情况分别为:

- (1) 情况1,旋转平面平行于卫星轨道面;
- (2) 情况2,旋转平面垂直于初始时刻卫星速度矢量;
- (3) 情况3,旋转平面垂直于初始时刻卫星位置矢量。

首先通过典型场景下的计算机计算结果,比较3种情况下的敏感度因子,然后结合这些结果进行分析。

场景参数:卫星轨道高度600km,信号频率3GHz,干涉仪基线长2m,干涉仪转速1转/10s,总观测时间10s,相位差变化率观测间隔0.5s,目标位于初始时刻星下点,即经、纬度均为0°。

3种情况下, η_L 和 η_B 随时间 t 变化曲线分别如图2和图3所示,相位差变化率随时间 t 变化曲线如图4所示。

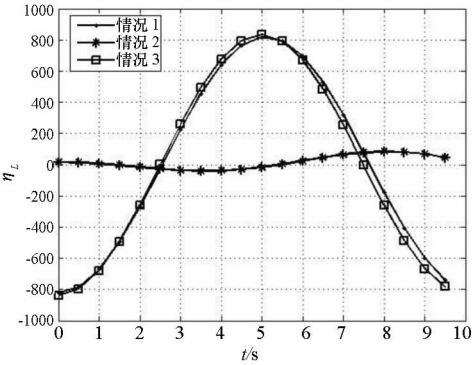


图2 η_L 随 t 变化曲线
Fig.2 Curve of η_L changing with time

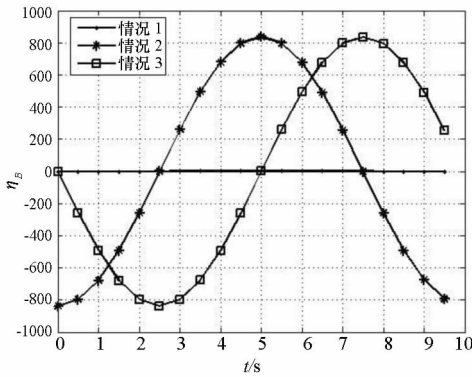


图3 η_B 随 t 变化曲线
Fig.3 Curve of η_B changing with time

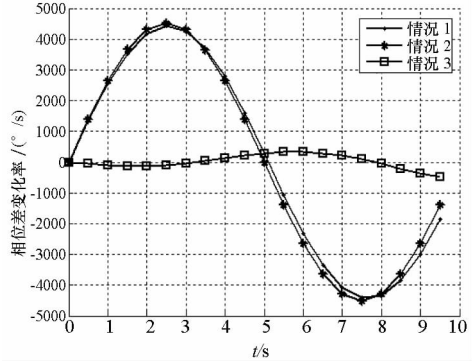


图4 相位差变化率随 t 变化曲线
Fig.4 Curve of $\dot{\varphi}$ changing with time

- 根据情况1,可以得到如下几条性质:
- 1. \mathbf{a}_L 平行于干涉仪旋转平面,且 $\|\mathbf{a}_L\| = R$;
 - 2. \mathbf{a}_B 垂直于干涉仪旋转平面,且 $\|\mathbf{a}_B\| = R$;
 - 3. 在初始观测时刻($m=0$), \mathbf{a}_L 垂直于 \mathbf{u}_0 ;
 - 4. 在整个观测过程中, \mathbf{u}_m 始终平行于干涉仪旋转平面,且 \mathbf{u}_m 的方向变化不大。
- 根据性质4, \mathbf{v}_m 始终平行于干涉仪旋转平面。根据性质1、3, \mathbf{v}_m 几乎平行于 \mathbf{a}_L ,当 $\mathbf{b}_{m\perp}$ 垂直于

u_m 时, $\|v_m\|$ 达到最大值 1, η_L 接近最大值 $\omega K \|a_L\|/r$ 。因此, η_L 近似按余弦规律变化, 振幅近似为 $\omega K \|a_L\|/r$, 如图 2 中情况 1 对应曲线, 对于 0° 经线上的其他目标, η_L 也具有这种变化规律, 振幅随目标纬度的增加而减小。

根据性质 2、4, v_m 始终垂直于 a_B , 所以 $|v_m^T a_B|$ 始终为 0, 另外, (20) 式中包含 ρ_m 的项也始终垂直于 a_B , 因此 η_B 始终为 0, 如图 3 中情况 1 对应曲线, 对于 0 度纬线上的其他目标, η_B 也始终为 0。

因此, 对于情况 1, 当目标处于星下点轨迹附近时, 定位误差较大, 特别地, 当目标处于星下点轨迹上时, 无法定位。

根据情况 2, 可以得到如下几条性质:

1. a_L 垂直于干涉仪旋转平面, 且 $\|a_L\| = R$;
2. a_B 平行于干涉仪旋转平面, 且 $\|a_B\| = R$;
3. 在初始观测时刻, a_B 垂直于 u_0 ;
4. 在初始观测时刻, u_0 平行于干涉仪旋转平面, 随着卫星的运动, u_m 与干涉仪旋转平面的夹角缓慢增大, 但是在观测持续的较短时间内, 该夹角始终较小。

经过与情况 1 类似的分析可得, η_L 近似按余弦规律变化, 振幅远小于 $\omega K \|a_L\|/r$, 如图 2 中情况 2 对应曲线, 对于 0 度经线上的其他目标, η_L 也具有这种变化规律, 只是振幅随着目标纬度的增加而减小; η_B 近似按余弦规律变化, 振幅近似为 $\omega K \|a_B\|/r$, 如图 3 中情况 2 对应曲线, 对于 0 度纬线上的其他目标, η_B 也具有这种变化规律, 振幅随目标经度的增加而减小。

因此, 对于情况 2, 当目标处于过初始时刻星下点的星下点轨迹垂线附近时, 定位误差较大。

根据情况 3, 可以得到如下几条性质:

1. a_L 平行于干涉仪旋转平面, 且 $\|a_L\| = R$;
2. a_B 平行于干涉仪旋转平面, 且 $\|a_B\| = R$;
3. 在初始观测时刻, u_0 垂直于干涉仪旋转平面, 随着卫星的运动, u_m 与干涉仪旋转平面不再严格垂直, 但是由于观测时间较短, 可以认为 u_m 与干涉仪旋转平面近似垂直。

根据与前两种情况类似的分析可得, η_L 近似按余弦规律变化, 振幅近似为 $\omega K \|a_L\|/r$, 如图 2 中情况 3 对应曲线, 对于 0 度经线上的其他目标, η_L 也具有这种变化规律, 振幅随着目标纬度的增加而减小; η_B 近似按余弦规律变化, 振幅近似为 $\omega K \|a_B\|/r$, 如图 3 中情况 3 对应曲线, 对于 0 度纬线上的其他目标, η_B 也具有这种变化规律, 振幅随目标经度的增加而减小。

因此, 对于情况 3, 不管是星下点轨迹附近的

目标还是过初始时刻星下点的星下点轨迹垂线附近的目标, η_L 和 η_B 都以较大的幅度近似按余弦规律变化, 对于星下点周围其他方向的目标, 也有类似规律, 因此定位误差在星下点周围各个方向的分布较为均匀, 不会出现在某个方向上定位误差很大或无法定位的情况。

由图 4 可以看出, 对于情况 3, 相位差变化率的量值及变化速度远小于情况 1 和情况 2, 这说明观测量本身的大小及变化快慢不能决定定位精度, 要综合考查观测量对于所有待估状态的可观测度, 因为如果观测量对于某一维状态的可观测度很强而对另一维状态的可观测度很弱, 那么就会出现观测量变化很快但定位精度很差的情况。

4 数值仿真与分析

4.1 定位误差计算

本节根据第 2 节推导得到的公式, 计算不同的干涉仪旋转平面以及转速情况下的 GDOP 分布, 对第 3 节的分析结果进行验证。

场景 1: 干涉仪旋转平面平行于卫星轨道面, 转速为 1 转/10s;

场景 2: 干涉仪旋转平面垂直于初始时刻卫星速度矢量, 转速为 1 转/10s;

场景 3: 干涉仪旋转平面垂直于初始时刻卫星位置矢量, 转速为 1 转/10s;

场景 4: 干涉仪旋转平面与场景 3 相同, 转速提高到 1 转/5s;

4 种场景中, 相位差变化率测量误差标准差均为 $10^\circ/\text{s}$, 其他参数同第 3 节。

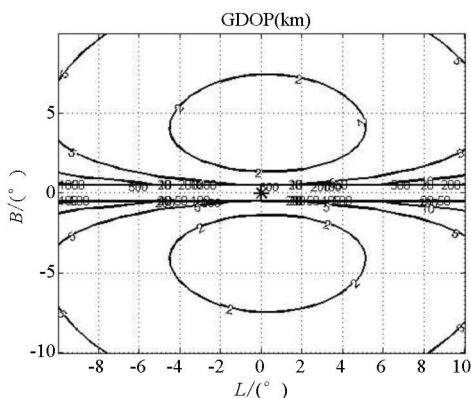


图 5 场景 1 GDOP

Fig. 5 GDOP of No. 1

图 5 ~ 图 8 分别给出了场景 1 ~ 场景 4 GDOP 分布图, 图中 “*” 表示初始观测时刻的星下点位置。

从图 5 可以看出, 当干涉仪旋转平面平行于卫星轨道面时, 对星下点轨迹正侧方目标的定位精度较好, 但是对星下点轨迹附近的目标定位精

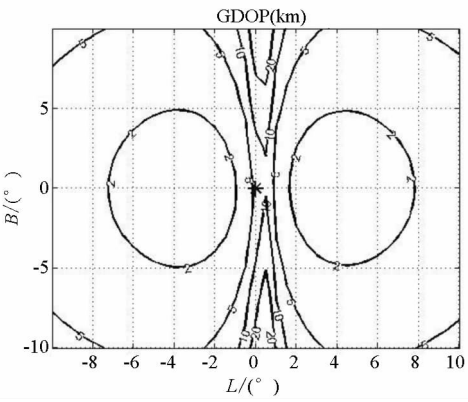


图6 场景2 GDOP
Fig. 6 GDOP of No. 2

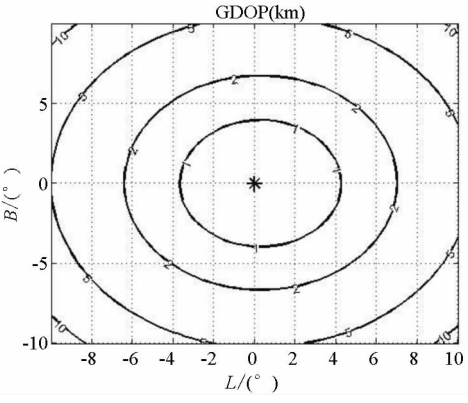


图7 场景3 GDOP
Fig. 7 GDOP of No. 3

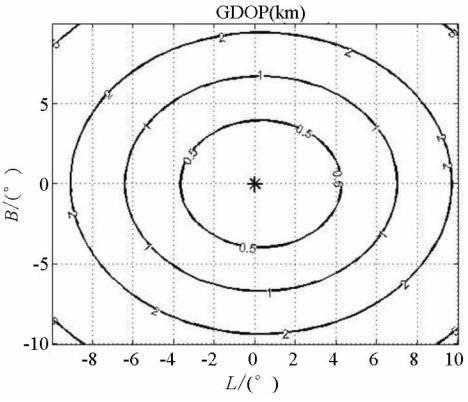


图8 场景4 GDOP
Fig. 8 GDOP of No. 4

度很差,这是因为此时 η_B 始终很小,对星下点轨迹上的目标则无法定位。

从图6可以看出,当干涉仪旋转平面垂直于初始时刻卫星速度矢量时,对星下点轨迹方向上的目标定位精度较好,但是对星下点轨迹正侧方目标的定位精度较差,这是因为此时 η_L 始终较小的原因。

从图7可以看出,当干涉仪旋转平面垂直于初始时刻卫星位置矢量时,定位误差分布比较均匀,对星下点各个方向的目标定位精度都较好,不存在定位精度很差或无法定位的区域,这是因为

这种情况下对于星下点各个方向的目标,都不会出现 η_L 或 η_B 始终很小的情况。

从图8与图7的比较中可以看出,在其他条件相同时,增大干涉仪转速,定位误差几乎成比例减小,这说明提高干涉仪转速能有效提高定位精度。

4.2 典型场景下的定位仿真

本节通过典型场景下的蒙特卡罗仿真,对基于PSO的定位算法的定位性能进行验证。根据前面的分析结果,当干涉仪旋转平面垂直于初始观测时刻卫星位置矢量时,定位误差分布比较均匀,总体定位性能最好,因此本节针对这种情况进行定位性能仿真。

目标位置设为距初始观测时刻星下点300km的地球表面上的一点,其他参数设置同第4.1节的情况。

在不同的相位差变化率测量误差情况下分别进行1000次蒙特卡罗仿真,统计得到定位均方根误差(RMSE)曲线以及对应的CRLB曲线如图9所示。

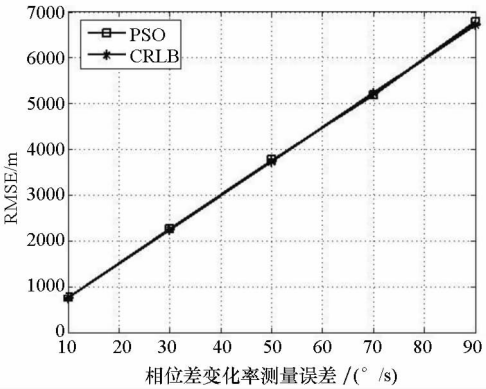


图9 不同测量误差下的定位误差
Fig. 9 Localization error under different noise levels

从以上仿真结果可以看出,在不同的测量误差水平下,基于PSO的定位算法均能够得到十分接近CRLB的定位精度,是本文所提方法的一种有效的定位求解算法。

5 总结与讨论

本文提出了一种在自旋卫星只安装两个接收通道构成一维干涉仪,通过测量干涉仪相位差变化率对地面固定辐射源定位的新方法,该方法所需设备量少,不受通道间幅度/相位不一致性的影响,能够实现快速高精度定位。理论分析和计算结果表明,当干涉仪旋转平面与初始观测时刻卫星位置矢量垂直时,定位误差的分布比较均匀,在星下点各个方向上都有较好的定位精度,增大干涉仪转速可以提高定位精度,基于PSO的定位算

法的定位精度接近 CRLB。

该定位方法本质上是利用干涉仪旋转带来的相位差变化信息进行定位,因此能否准确测量到该变化信息将直接影响定位精度。在实际中,相位差变化率一般通过多个脉冲的相位差拟合得到,当干涉仪转动速度较低时,可以通过线性或二阶模型得到精度较高的相位差变化率参数,但是当干涉仪转速较高时,相位差剧烈变化,此时必须用更高阶的模型来描述相位差的变化规律,相位差的估计精度也会随之下降。因此在给定的条件下,应该存在一个对应的干涉仪转速上限,低于该上限时,提高干涉仪转速有利于提高定位精度;高于该上限时,增加干涉仪转速将无益于提高定位精度或者由于相位差变化率难以估计而导致定位精度下降。另外,由于无法得到角度信息,所以信号分选方面还需要研究新的技术。关于这些问题的研究将逐步展开。

参考文献 (References)

- [1] 徐义, 郭福成, 冯道旺. 一种单星仅测 TOA 无源定位方法[J]. 宇航学报, 2010, 31(2): 502–508.
XU Yi, GUO Fucheng, FENG Daowang. A new satellite passive localization method using TOA measurement only[J]. Journal of Astronautics, 2010, 31(2): 502–508. (in Chinese)
- [2] 陆安南, 孔宪正. 单星测频无源定位法[J]. 通信学报, 2004, 25(9): 160–168.
LU Annan, KONG Xianzheng. Passive localization from frequency measurements by single satellite[J]. Journal on Communications, 2004, 25(9): 160–168. (in Chinese)
- [3] 张敏, 冯道旺, 郭福成. 基于多普勒变化率的单星无源定位[J]. 航天电子对抗, 2009, 25(5): 11–13.
ZHANG Min, FENG Daowang, GUO Fucheng. Passive localization by a single satellite based on doppler rate-of-change[J]. Aerospace Electronic Counter-measurement, 2009, 25(5): 11–13. (in Chinese)
- [4] 钟丹星. LBI 体制运动单平台无源定位关键技术研究[D]. 长沙: 国防科技大学, 2007.
ZHONG Danxing. Research on key technologies of LBI passive location system boarded on single moving observer[D]. Changsha: National University of Defense Technology, 2007. (in Chinese)
- [5] 刘海军, 柳征, 姜文利, 等. 基于星载测向体制的辐射源定位融合算法[J]. 系统工程与电子技术, 2009, 31(12): 2875–2878.
LIU Haijun, LIU Zheng, JIANG Wenli, et al. Integration algorithm of emitter position based on satellites direction-finding systems[J]. Systems Engineering and Electronics, 2009, 31(12): 2875–2878. (in Chinese)
- [6] 袁孝康. 相位干涉仪测向定位研究[J]. 上海航天, 1999(3): 1–7.
YUAN Xiaokang. Study on direction-finding and position with phase interferometers[J]. Aerospace Shanghai, 1999(3): 1–7. (in Chinese)
- [7] Jacobs E, Ralston E W. Ambiguity resolution in interferometry[J]. IEEE Transactions on Aerospace and Electronic Systems, 1981, 17(6): 766–780.
- [8] 周亚强, 皇甫堪. 噪声条件下数字式多基线相位干涉仪解模糊问题[J]. 通信学报, 2005, 26(8): 16–21.
ZHOU Yaqiang, HUANGFU Kan. Solving ambiguity problem of digitized multi-baseline interferometer under noisy circumstance[J]. Journal on Communications, 2005, 26(8): 16–21. (in Chinese)
- [9] 张文旭, 司锡才, 蒋伊琳. 相位干涉仪测向系统相位误差研究[J]. 系统工程与电子技术, 2006, 28(11): 1631–1633.
ZHANG Wenxu, SI Xicai, JIANG Yilin. Research on phase error in the system of direction finding by phase interferometer[J]. System Engineering and Electronics, 2006, 28(11): 1631–1633. (in Chinese)
- [10] 郭福成, 贾兴江, 皇甫堪. 仅用相位差变化率的机载单站无源定位方法及其误差分析[J]. 航空学报, 2009, 30(6): 1090–1095.
GUO Fucheng, JIA Xingjiang, HUANGFU Kan. A single observer passive localization method using phase difference changing rate only and its error analysis[J]. Acta Aeronautica Et Astronautica Sinica, 2009, 30(6): 1090–1095. (in Chinese)
- [11] Wang J H, Wang Y S, GUO T, et al. Rate of phase difference change estimation in single airborne passive locating system[J]. Chinese Journal of Aeronautics, 2009, 22(2): 184–190.
- [12] 章仁为. 卫星轨道姿态动力学与控制[M]. 北京: 北京航空航天大学出版社, 1998: 139.
ZHANG Renwei. Satellite orbit attitude kinetics and control[M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 1998: 139. (in Chinese)
- [13] 解可新, 韩健, 林友联. 最优化方法(修订版)[M]. 天津: 天津大学出版社, 2004.
XIE Kexin, HAN Jian, LIN Youlian. Optimization method (revised edition)[M]. Tianjin: Tianjin University Press, 2004. (in Chinese)
- [14] 郭福成. 基于多级网格搜索的固定单站无源定位方法[J]. 信号处理, 2008, 24(6): 927–930.
GUO Fucheng. A fixed observer passive location method based on multi-level grid search[J]. Signal Processing, 2008, 24(6): 927–930. (in Chinese)
- [15] Kennedy J, Eberhart R. Particle swarm optimization[C]// Proceedings of the IEEE International Conference on Neural Networks, 1995: 1942–1948.
- [16] Shi Y H, Eberhart R C. A modified particle swarm optimizer[C]// Proceedings of the IEEE International Conference on Evolutionary Computation, 1998: 69–73.
- [17] Seo J H, Im C H, Heo C G, et al. Multimodal function optimization based on particle swarm optimization[J]. IEEE Transactions on Magnetics, 2006, 42(4): 1095–1098.
- [18] 王芳. 粒子群算法的研究[D]. 重庆: 西南大学, 2006.
WANG Fang. Research on particle swarm algorithms[D]. Chongqing: Southwest University, 2006. (in Chinese)
- [19] 孙仲康, 周一字, 何黎星. 单多基地有源无源定位技术[M]. 北京: 国防工业出版社, 1996: 37.
SUN Zhongkang, ZHOU Yiyu, HE Lixing. Mono-station/multi-station passive/active location technology[M]. Beijing: National Defense Industry Press, 1996: 37. (in Chinese)
- [20] Kalaba R E, Spingarn K. Optimal inputs for nonlinear process parameter estimation[J]. IEEE Transactions on Aerospace and Electronic Systems, 1974, 10(3): 339–345.