基于正交梯度搜索的动态系统递阶优化辨识

衷路生 [1,2] 宋执环 [1]

摘 要 提出了一种辨识线性时不变状态空间系统参数的正交梯度二步递阶优化方法. 通过极小化输出误差目标函数获得了系统参数估计; 提出了正交梯度搜索方法用于解决系统参数的非唯一性问题, 正交梯度搜索的本质是在输入 - 输出等价类相切平面的正交垂空间更新系统参数; 给出了用 L-M 算法进行参数优化的充分条件; 提出的系统参数递阶优化辨识方法包括两步: 首先用给出的自适应 L-M 算子正交梯度方法确定参数优化方向; 其次由一维搜索方法计算最佳步长. 蒙特卡罗数值仿真实验表明本文提出的方法具有收敛速度快、抗噪能力强以及数值稳定性好等优点.

关键词 系统辨识, 参数估计, Levenberg-Marquardt (L-M) 算法, 递阶优化, 状态空间模型, 极大似然估计
中图分类号 TP273

Hierarchical Optimization Identification of LTI State-space Systems by Projected Gradient Search

ZHONG Lu-Sheng[1,2] SONG Zhi-Huan[1]

Abstract A hierarchical optimization method for the identification of LTI state-space systems is proposed based on projected gradient search. The system parameters are determined by optimizing an output-error cost function. To deal with the non-uniqueness of the fully parameterized state-space system, a projected gradient search algorithm is presented by restricting the update of the parameters to the tangent space to the manifold of observationally equivalent state-space systems. The sufficient condition to employ L-M algorithm for optimizing parameters is also introduced. The proposed hierarchical optimization identification method includes two steps: First, the parameter search direction is determined by the proposed adaptive L-M projected gradient approach; Second, the optimum step size is computed according to a line search method. Numerical simulation studies show that the proposed algorithm offers improved performance, such as faster convergence speed and better numerical stability, over existing EM and DDLC methods.

Key words System identification, parameter estimation, Levenberg-Marquardt (L-M) algorithm, hierarchical optimization (HO), state-space models, maximum likelihood estimation

由输入输出观测数据获得多输入多输出线性时不变状态空间系统的参数估计是系统分析、设计和控制的关键环节[1], 长期以来一直作为热点问题而受到广泛研究. 许多辨识方法也相继被提出用以估计多输入多输出状态空间系统的参数, 这些方法依据系统参数的更新方式可以分为两大类: 非迭代方法和迭代方法, 这两类方法的特点归纳如下:

Verhaegen, Overschee, Qin 等研究人员提出的状态空间系统子空间辨识方法[1−4] 属于参数非迭代估计范畴. 子空间方法的核心思想是, 将状态空间系统的输入输出数据组成的观测数据矩阵进行空间投影运算, 投影结果的奇异值分解得到的分块子矩阵包含了状态空间系统的扩展能控性矩阵以及非平稳卡尔曼状态估计[2] 等信息, 利用这些信息便可得到状态空间系统的系统参数及噪声参数的估计. 子空间方法的优点是系统参数、噪声参数以及状态序列的估计一步即可完成, 无需反复迭代计算. 这种方法的不足之处是: 1) 子空间方法不是按照某个最优准则求解系统参数, 因而只能得到 "次优" 的估计结果, 其参数估计精度一般不如极大似然方法或预报误差方法[4]; 2) 由于估计误差的影响, 子空间方法得到的估计模型可能不满足正实性条件, 此时子空间方法将失效[5−6].

状态空间系统参数辨识的另一个重要分支是参数迭代估计方法[7−14]. Mckelvey, Ribarits 的 DDLC (Data driven local coordinates) 方法[7−10], Bergboer, Gibson 的极大似然辨识方法[11], Verdult 等的梯度搜索方法[12−14] 等都属于迭代学习的范畴. Mckelvey, Ribarits 的 DDLC 方法[7−10], 其不足之处是无法确定系统最小实现的参数空间集; Gibson 等利用极大期望原理, 提出了由输入输出数据构成的条件概率意义下的状态空间系统参数的极大似然辨识方法[11]. 这种方法具有许多优点, 如: 1) 每步迭代中得到的参数更新值可以确保系统的似然函数值的单调非递减性; 2) 对于能控能观的状态空间系统, 该算法能够使系统参数从选定的初始值以指数收敛的速度经过有限次迭代达到稳定点. Gibson 提出的方法不足之处是对每个观测数据都要计算卡尔曼滤波以及定区间平滑的系统状态估计向量、状态方差矩阵, 在观测数据很大以及系统维数很高时将耗费大量的计算机存储空间, 影响数据处理速度. 此外, Bergboer, Verdult 等的梯度搜索方法[12−14] 由于其 L-M 步长因子的取值缺乏理论依据, 通常是设为定常值, 由此带来了参数收敛速度慢、参数容易陷入局部极小甚至发散等问题.

基于以上各种线性时不变状态空间系统参数辨识方法的分析, 本文提出基于正交梯度搜索的状态空间系统递阶优化辨识方法. 二步递阶优化包括了在正交垂平面上确定系统参数优化方向以及计算最佳步长. 数值仿真实验结果表明本文提出的方法具有收敛速度快、抗噪能力强以及数值稳定性好等优点.

1 辨识问题描述

考虑如下线性时不变离散状态空间系统

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + B\boldsymbol{u}_k + G\boldsymbol{v}_k \tag{1}$$

$$\boldsymbol{y}_k = C\boldsymbol{x}_k + D\boldsymbol{u}_k + \boldsymbol{v}_k \tag{2}$$

其中, $\boldsymbol{x}_k$ 是 n 维状态向量, $\boldsymbol{y}_k$ 是 l 维输出观测向量, $\boldsymbol{u}_k$ 是 m 维输入观测向量, 系统参数矩阵 $A \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times m}$, $C \in \mathbf{R}^{l\times n}$, $D \in \mathbf{R}^{l\times m}$, $G \in \mathbf{R}^{n\times l}$, $\boldsymbol{v}_k$ 是 l 维测量噪声向量, 服从如下高斯分布

$$\boldsymbol{v}_k \sim N(0, \sigma^2 I_l) \tag{3}$$

其中, $\sigma^2 > 0 \in \mathbf{R}$, I_l 表示 $l \times l$ 单位矩阵.

为分析方便, 定义参数向量

$$\boldsymbol{\theta} = [vec(A)^{\mathrm{T}}, vec(\hat{B})^{\mathrm{T}}, vec(C)^{\mathrm{T}}, vec(D)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbf{R}^d \tag{4}$$

收稿日期 2006-11-30 收修改稿日期 2007-04-10
Received November 30, 2006; in revised form April 10, 2007
国家重点基础研究发展计划 (973 计划) (2002CB312203-02), 浙江省自然科学基金项目 (601114) 资助
Supported by National Basic Research Program of China (973 Program) (2002CB312203-02) and Natural Science Foundation of Zhejiang Province (601114)
1. 浙江大学工业控制研究所工业控制技术国家重点实验室 杭州 310027 2. 华东交通大学电气与电子工程学院 南昌 330013
1. State Key Laboratory of Industrial Control Technology, Institution of Industrial Process Control, Zhejiang University, Hangzhou 310027 2. School of Electrical and Electronic Engineering, East China Jiaotong University, Nanchang 330013
DOI: 10.3724/SP.J.1004.2008.00711

其中, $\hat{B}=(B,G)$, $d=n^2+n(m+2l)+lm$ 是系统参数总个数, 符号 $vec(\Delta)$ 表示对矩阵 Δ 进行拉直运算, Δ^{T} 是转置符号.

辨识问题可以表述为: 给定输入 - 输出观测序列 (U_N, Y_N), 其中

$$\begin{aligned} U_N &= \{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_{N-1},\boldsymbol{u}_N\} \\ Y_N &= \{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_{N-1},\boldsymbol{y}_N\} \end{aligned} \tag{5}$$

估计参数矩阵 A,B,C,D,G, 使得以下目标函数达到最小

$$J_N(\boldsymbol{\theta})=\sum_{k=1}^{N}\|\boldsymbol{y}_k-\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta})\|_2^2=E_N^{\mathrm{T}}(\boldsymbol{\theta})E_N(\boldsymbol{\theta}) \tag{6}$$

其中, $\boldsymbol{e}_k(\boldsymbol{\theta})=\boldsymbol{y}_k-\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta}), E_N(\boldsymbol{\theta})=[\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{\theta}),\cdots,\boldsymbol{e}_N^{\mathrm{T}}(\boldsymbol{\theta})]^{\mathrm{T}}$, 并且假定目标函数 $J_N(\boldsymbol{\theta})$ 对任意的参数 $\boldsymbol{\theta}_i\,(i=1,2,\cdots,d)$ 都二次连续可微, $\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta})$ 表示基于 $(k-1)$ 时刻对 k 时刻系统输出的一步超前预报, 可以由以下均方一步最优估计器计算

$$\begin{aligned} \hat{\boldsymbol{x}}_{k+1|k} &= (A-GC)\hat{\boldsymbol{x}}_{k|k-1}+(B-GD)\boldsymbol{u}_k+G\boldsymbol{y}_k= \\ &\bar{A}\hat{\boldsymbol{x}}_{k|k-1}+\bar{B}\boldsymbol{u}_k+G\boldsymbol{y}_k \end{aligned} \tag{7}$$

$$\hat{\boldsymbol{y}}_{k|k-1}=C\hat{\boldsymbol{x}}_{k|k-1}+D\boldsymbol{u}_k \tag{8}$$

其中, $\bar{A}=A-GC$, $\bar{B}=B-GD$; 如果动态系统 (1) 和 (2) 能控能观, 并且矩阵 $\bar{A}=A-GC$ 和 A 的最大特征值满足 $\lambda_1=\max|\lambda(\bar{A})|<1$, $\lambda_2=\max|\lambda(A)|<1$, 则由式 (7) 和 (8) 可以得到状态以及输出的一致渐近估计.

类似文献 [11] 的分析, 当噪声服从式 (3) 的高斯分布时, 极小化目标函数 (6) 得到的参数估计等价于动态系统 (1) 和 (2) 的参数极大似然估计.

2 正交梯度分析

本节分析了基于一阶泰勒展开公式的目标函数近似优化原理, 根据随机状态空间相似变换原理以及矩阵摄动分析得到了输入 - 输出等价类及其相切平面的数学表达式, 进而提出了在正交垂平面优化系统参数的方法.

2.1 输入输出等价类及其相切平面

由线性系统理论可知, 动态系统 (1) 和 (2) 在输入 - 输出观测序列 (U_N,Y_N) 上等价于参数 $\tilde{\boldsymbol{\theta}}=[vec(\tilde{A})^{\mathrm{T}},vec(\tilde{B})^{\mathrm{T}},vec(\tilde{C})^{\mathrm{T}},vec(\tilde{D})^{\mathrm{T}}]^{\mathrm{T}}$ 确定的动态系统 (也称 $\tilde{\boldsymbol{\theta}}$ 和 $\boldsymbol{\theta}$ 满足相似变换或输入输出等价类), 其中

$$\tilde{A}=TAT^{-1},\ \tilde{B}=T\hat{B},\ \tilde{C}=CT^{-1},\ \tilde{D}=D \tag{9}$$

$T\in\mathbf{R}^{n\times n}$ 为非奇异矩阵.

接下来利用局部线性化方法来找出输入 - 输出等价类 $\tilde{\boldsymbol{\theta}}$ 和 $\boldsymbol{\theta}$ 所满足的数学关系. 相似变换矩阵取单位阵 I_n 与微小摄动矩阵 ΔT 之和, 即: $T=I_n+\Delta T$, 由 $(I_n+\Delta T)^{-1}=I_n-\Delta T$, 并且忽略 ΔT 的高阶项可得

$$\begin{aligned} \tilde{\boldsymbol{\theta}} &= [vec(\tilde{A})^{\mathrm{T}},vec(\tilde{B})^{\mathrm{T}},vec(\tilde{C})^{\mathrm{T}},vec(\tilde{D})^{\mathrm{T}}]^{\mathrm{T}}= \\ &[vec(TAT^{-1})^{\mathrm{T}},vec(T\hat{B})^{\mathrm{T}},vec(CT^{-1})^{\mathrm{T}},vec(D)^{\mathrm{T}}]^{\mathrm{T}} \end{aligned}$$

也即

$$\tilde{\boldsymbol{\theta}}=\begin{pmatrix} vec(A)\\ vec(\hat{B})\\ vec(C)\\ vec(D)\end{pmatrix}+\underbrace{\begin{pmatrix} vec(-A\Delta T+\Delta TA)\\ vec(\Delta T\hat{B})\\ vec(-C\Delta T)\\ vec(0_{l\times m})\end{pmatrix}}_{term1} \tag{10}$$

利用矩阵的拉直公式[7] $vec(ABC)=(C^{\mathrm{T}}\otimes A)vec(B)$, 则式 (10) 中的 $term1$ 可以写成

$$term1=\underbrace{\begin{pmatrix} -I_n\otimes A+A^{\mathrm{T}}\otimes I_n\\ \hat{B}^{\mathrm{T}}\otimes I_n\\ -I_n\otimes C\\ 0_{lm\times n^2}\end{pmatrix}}_{Q}vec(\Delta T) \tag{11}$$

矩阵 Q 也称为输入 - 输出等价类的相切平面. 由线性系统理论可知, 如果系统 (A,C) 能观测并且 $(A,\hat{B})$ 可控, 则矩阵 Q 的秩等于 n^2.

也就是说, 式 (6) 的目标函数沿矩阵 Q 的列向量张成的搜索方向上的取值将保持不变. 为了克服输入 - 输出等价类对目标函数的取值不变性影响, 应该在输入 - 输出等价类相切平面的正交垂平面方向 $Q^{\perp}$ 上优化系统参数. 矩阵 $Q^{\perp}$ 可以通过相切平面奇导值分解[3] 计算得到.

2.2 正交垂平面上的目标函数表示

记目标函数式 (6) 的沿参数正交垂平面 $Q^{\perp}$ 方向上优化参数 $\boldsymbol{\theta}^{\perp}$ 的维数为 $d_{\perp}$, 则

$$d_{\perp}=\mathrm{rank}\,(Q^{\perp})=d-r \tag{12}$$

依据多元函数的一阶泰勒近似原理, 在参数 $\boldsymbol{\theta}^{\perp}$ 下极小化目标函数 (6) 等价于在参数正交垂平面 $Q^{\perp}$ 方向上优化以下表达式

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{k+1}^{\perp} &= \arg\min_{\boldsymbol{\theta}^{\perp}}J_N(\boldsymbol{\theta}^{\perp})\cong\arg\min_{\boldsymbol{\theta}^{\perp}}J_N(\hat{\boldsymbol{\theta}}_k^{\perp})= \\ &E_N^{\mathrm{T}}(\hat{\boldsymbol{\theta}}_k)E_N(\hat{\boldsymbol{\theta}}_k)+(g|_{\boldsymbol{\theta}_k^{\perp}})^{\mathrm{T}}(\boldsymbol{\theta}^{\perp}-\hat{\boldsymbol{\theta}}_k^{\perp})+ \\ &\frac{1}{2}(\boldsymbol{\theta}^{\perp}-\hat{\boldsymbol{\theta}}_k^{\perp})^{\mathrm{T}}(T|_{\boldsymbol{\theta}_k^{\perp}})(\boldsymbol{\theta}^{\perp}-\hat{\boldsymbol{\theta}}_k^{\perp}) \end{aligned} \tag{13}$$

其中

$$\begin{aligned} g|_{\boldsymbol{\theta}_k^{\perp}} &= \nabla J_N(\boldsymbol{\theta}^{\perp})=2R(\boldsymbol{\theta}^{\perp})^{\mathrm{T}}E_N(\boldsymbol{\theta}) \\ T|_{\boldsymbol{\theta}_k^{\perp}} &= 2R(\boldsymbol{\theta}^{\perp})^{\mathrm{T}}R(\boldsymbol{\theta}^{\perp}) \end{aligned} \tag{14}$$

$R(\boldsymbol{\theta}^{\perp})$ 为输出误差对优化参数 $\boldsymbol{\theta}^{\perp}$ 的 Jacobian 矩阵. 其定义式为

$$R(\boldsymbol{\theta}^{\perp})_{i,j}=\frac{\partial e_i(\theta)}{\partial\theta_j^{\perp}}\ (i=1,\cdots,N,\ j=1,\cdots,d_{\perp}) \tag{15}$$

在第 $k+1$ 步, 系统原参数和正交垂平面方向的优化参数的关系式为

$$\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k+Q_k^{\perp}(\boldsymbol{\theta}_{k+1}^{\perp}-\boldsymbol{\theta}_k^{\perp}) \tag{16}$$

由式 (13) ~ (15) 可知, 式 (13) 的优化问题最关键的是计算出 $R(\boldsymbol{\theta}^{\perp})$, 本文采用沿正交垂平面 $Q^{\perp}$ 方向上的导数逼近方法来计算 $R(\boldsymbol{\theta}^{\perp})$, 具体做法是: 假设 $Q_{(:,i)}^{\perp}$ 表示矩阵 $Q^{\perp}$ 的第 i 列, $R(\boldsymbol{\theta}^{\perp})_{(:,i)}$ 表示矩阵 $R(\boldsymbol{\theta}^{\perp})$ 的第 i 列 $(i=1,2,\cdots,d_{\perp})$, 则输出误差向量 $E_N(\boldsymbol{\theta})$ 对正交垂平面 $Q^{\perp}$ 方向上参数 $\boldsymbol{\theta}_i^{\perp}$ 的导数的数值逼近式为

$$R(\boldsymbol{\theta}^{\perp})_{(:,i)}=\lim_{\gamma\to 0}\begin{pmatrix} \dfrac{e_1(\boldsymbol{\theta}+Q_{(:,i)}^{\perp}\gamma)-e_1(\boldsymbol{\theta})}{\gamma}\\ \cdots\\ \cdots\\ \dfrac{e_N(\boldsymbol{\theta}+Q_{(:,i)}^{\perp}\gamma)-e_N(\boldsymbol{\theta})}{\gamma}\end{pmatrix} \tag{17}$$

3 状态空间系统参数的正交梯度二步递阶优化辨识方法

极小化目标函数 (6)、(13) 是一个非线性最小二乘参数问题, 对于这一类问题的求解方法, Bergboer[12], Verdult[13] 等学者在梯度学习中加入正则化因子, 并且利用矩阵的 QR 分解原理避免高维矩阵的直接求逆, 在一定程度上提高了算法的稳定性, 但却带来了参数收敛速度慢, 算法效率低的不足; Ribarits 提出的参数局部调整方法[8−10] 只能得到使系统达到最小实现的参数.

本文提出正交梯度的二步递阶优化方法来辨识状态空间系统参数. 第一步, 在正交梯度原理基础上, 用自适应步长因子的 Levenberg-Marquardt 算法得到参数的更新方向; 第二步, 由一维搜索方法计算系统参数沿更新方向的最佳步长. 第一步保证了算法的稳定性, 第二步则能加快系统参数的收敛速度.

3.1 确定参数搜索方向以及计算最佳步长

Levenberg-Marquardt 算法是处理非线性优化问题的有力工具. 和一般的梯度算法相比, Levenberg-Marquardt 算法具有收敛速度快、稳定性好等优点, 因此, 在神经网络学习[15]、非线性学习控制[16], 信号处理[17] 等领域得到了广泛应用. 使用固定的 L-M 算子是文献 [15−17] 的共同点, 由此带来的不足是目标函数下降速度慢. 为此, 本文采取自适应步长因子的 L-M 算法来加快参数学习的速度, 提高算法的效率.

首先介绍状态空间系统基于 Levenberg-Marquardt 算法的正交梯度参数估计原理, 如定理 1 所述.

定理 1. 若动态系统 (1) 和 (2) 满足如下条件:

1) 系统满足 $\{A,(B,G)\}$ 能控, $\{A,C\}$ 能观测;

2) 输入序列 U_N 有界;

3) 由式 (12) 定义的 $T|_{\boldsymbol{\theta}_k^\perp}$ 恒为正定矩阵.

则对任意的 $\beta\in(0,1)$, 都存在着 $\tilde{\alpha}>0$, 使得只要 $\alpha>\tilde{\alpha}$, 对于可行域中任意的参数 $\boldsymbol{\theta}$ 都有

$$\begin{aligned}&J_N(\boldsymbol{\theta}+Q^\perp p(\alpha,\boldsymbol{\theta}^\perp))\le\\&J_N(\boldsymbol{\theta})+2\beta(p(\alpha,\boldsymbol{\theta}^\perp))^{\mathrm{T}}(Q^\perp)^{\mathrm{T}}((Q^\perp)^{\mathrm{T}})^{+}g|_{\boldsymbol{\theta}}^\perp\end{aligned}\tag{18}$$

其中, $((Q^\perp)^{\mathrm{T}})^{+}=(Q^\perp(Q^\perp)^{\mathrm{T}})^{-1}Q^\perp$ 是 Moore-Penrose 广义逆, $p(\alpha,\boldsymbol{\theta}^\perp)$ 是式 (19) 的解

$$(T|_{\boldsymbol{\theta}}^\perp+\alpha W(\boldsymbol{\theta}^\perp))p(\alpha,\boldsymbol{\theta}^\perp)=-g|_{\boldsymbol{\theta}}^\perp\tag{19}$$

式 (19) 中, $W(\boldsymbol{\theta}^\perp)$ 为一连续的正定对角阵, 在实际应用中 $W(\boldsymbol{\theta}^\perp)$ 常取单位阵, α 称为 Levenberg-Marquardt 算子, $g|_{\boldsymbol{\theta}_k^\perp}$, $T|_{\boldsymbol{\theta}_k^\perp}$ 的定义见式 (14).

证明. 条件 1) 和 2) 保证了系统 (1) 和 (2) 的观测输出有界, 进而确保了 $g|_{\boldsymbol{\theta}_k^\perp}, T|_{\boldsymbol{\theta}_k^\perp}$ 的存在性. 将状态空间系统全参数形式的 Levenberg-Marquardt 算法结论[18] 应用于正交垂平面的参数优化, 即可证明定理 1, 从略. □

依据式 (19), 定理 1 说明, 当式 (6) 的目标函数 $J_N(\boldsymbol{\theta})$ 还没达到最小值并且优化问题在参数垂平面方向上进行, 适当选择 α 并以式 (19) 的解 $p(\alpha,\boldsymbol{\theta}^\perp)$ 作为增量, 总能使目标函数 $J_N(\boldsymbol{\theta})$ 值获得一定的下降量.

在完成第一步, 即确定参数更新的方向 $Q^\perp p(\alpha,\boldsymbol{\theta}^\perp)$ 后, 还要计算出沿这个方向使目标函数下降最多的最佳步长, 即求解

$$\begin{aligned}&\arg\min_\lambda J_N(\boldsymbol{\theta}+\lambda Q^\perp p(\alpha,\boldsymbol{\theta}^\perp))=\\&\arg\min_\lambda\sum_{k=1}^{N}\|\boldsymbol{y}_k-\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta}+\lambda Q^\perp p(\alpha,\boldsymbol{\theta}^\perp))\|_2^2\end{aligned}\tag{20}$$

在优化理论中, 求解式 (20) 称为一维搜索, 常用的方法有: Fibonacci 方法、0.618 法、抛物线插值法、切线法和二分法等. 由于 0.618 法具有搜索效率高、计算方便等优点, 因此本文采用 0.618 法来计算最佳步长.

3.2 辨识算法小结

本节给出了状态空间系统参数的正交梯度二步递阶优化思想的辨识算法. 以下是依据输出误差实现自适应 L-M 算子的具体步骤:

Step 1. 选取正定对角矩阵 $W=I$, 同时选取 $\beta\in(0,1)$, 设定梯度范数阈值 ξ 和近似导数步长 γ, 设定自适应 L-M 步长因子初始值 $\alpha_0>0$;

Step 2. 设定系统输出误差和步长因子对 $(\Delta E_1,\eta_1)$ $(\Delta E_2,\eta_2)$, $\cdots$, $(\Delta E_{q-1},\eta_{q-1})(\Delta E_q,\eta_q)$, 满足关系: $\Delta E_1>\Delta E_2>\cdots>\Delta E_{q-1}>\Delta E_q$, ΔE_1 通常取系统输出初始估计误差, $\eta_1>\eta_2>\cdots>\eta_{q-1}>\eta_q>1$, q 称为自适应步长因子等级数;

Step 3. 设定系统参数初始值 $\hat{\boldsymbol{\theta}}_0$, 使之满足定理 1 各个条件, 并记迭代次数标记为 $j=0$, 置 $M=\eta_1$, $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0$;

Step 4. 依据式 (18) 计算输入 - 输出等价类相切平面 Q, 由奇异值分解得到正交垂平面 $Q^\perp$, 由式 (7) 和 (8) 分别计算 $\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta})$, $\hat{\boldsymbol{y}}_{k|k-1}(\boldsymbol{\theta}+Q_{(:,i)}^\perp\gamma)$, 得到误差向量 $E_N(\boldsymbol{\theta}+Q_{(:,i)}^\perp\gamma)=[\boldsymbol{y}_1^{\mathrm{T}}-\hat{\boldsymbol{y}}_{1|0}^{\mathrm{T}}(\boldsymbol{\theta}+Q_{(:,i)}^\perp\gamma),\cdots,\boldsymbol{y}_N^{\mathrm{T}}-\hat{\boldsymbol{y}}_{N|N-1}^{\mathrm{T}}(\boldsymbol{\theta}+Q_{(:,i)}^\perp\gamma)]^{\mathrm{T}}$, $E_N(\boldsymbol{\theta})=[\boldsymbol{y}_1^{\mathrm{T}}-\hat{\boldsymbol{y}}_{1|0}^{\mathrm{T}}(\boldsymbol{\theta}),\cdots,\boldsymbol{y}_N^{\mathrm{T}}-\hat{\boldsymbol{y}}_{N|N-1}^{\mathrm{T}}(\boldsymbol{\theta})]^{\mathrm{T}}$ $(k=1,2,\cdots,N$, $i=1,2,\cdots,d_\perp)$, 根据式 (17) 计算 $R(\boldsymbol{\theta}^\perp)$, 同时计算 $g|_{\boldsymbol{\theta}^\perp}=\nabla J_N(\boldsymbol{\theta}^\perp)=2R(\boldsymbol{\theta}^\perp)^{\mathrm{T}}E_N(\boldsymbol{\theta})$, 如果 $\|Q^\perp g|_{\boldsymbol{\theta}^\perp}\|<\xi$, 则停止计算;

Step 5. 计算 $T|_{\boldsymbol{\theta}^\perp}=2R(\boldsymbol{\theta}^\perp)^{\mathrm{T}}R(\boldsymbol{\theta}^\perp)$, 求解方程

$$(T|_{\boldsymbol{\theta}^\perp}+\alpha_jW)p(\alpha_j,\boldsymbol{\theta}^\perp)=-g|_{\boldsymbol{\theta}^\perp}\tag{21}$$

Step 6. 检验条件 $J_N(\boldsymbol{\theta}+Q^\perp p(\alpha_j,\boldsymbol{\theta}^\perp))\le J_N(\boldsymbol{\theta})+2\beta(p(\alpha_j,\boldsymbol{\theta}^\perp))^{\mathrm{T}}(Q^\perp)^{\mathrm{T}}((Q^\perp)^{\mathrm{T}})^{+}g|_{\boldsymbol{\theta}^\perp}$ 是否成立, 如果不成立, 则置 $\alpha_j=M\alpha_j$, 转 Step 5, 继续求解方程 (21) 得 $p(\alpha_j,\boldsymbol{\theta}^\perp)$; 否则由一维搜索 (0.618 法) 求解最佳步长 $\lambda^*=\arg\min_\lambda J_N(\boldsymbol{\theta}+\lambda Q^\perp p(\alpha_j,\boldsymbol{\theta}^\perp))$, 置 $\boldsymbol{\theta}=\boldsymbol{\theta}+\lambda^*Q^\perp p(\alpha_j,\boldsymbol{\theta}^\perp))$, 依据式 (6) 计算在新参数 $\boldsymbol{\theta}$ 下的目标函数值 $J_N(\boldsymbol{\theta})$, 遍历 $t=1,2,\cdots,q$, 确定 $J_N(\boldsymbol{\theta})\in[\Delta E_{t-1},\Delta E_t]$, 置 $\alpha_{j+1}=\alpha_j/\eta_{t-1}$, $M=\eta_{t-1}$;

Step 7. 置 $j=j+1$, 转 Step 4.

4 仿真与结果分析

本节应用前面介绍的辨识算法对状态空间系统进行数值仿真. 考虑由文献 [2] 给出的状态空间模型

$$\begin{aligned}\boldsymbol{x}_{k+1}&=A\boldsymbol{x}_k+B\boldsymbol{u}_k+K\boldsymbol{e}_k\\\boldsymbol{y}_k&=C\boldsymbol{x}_k+D\boldsymbol{u}_k+0.9736\,\boldsymbol{e}_k\end{aligned}$$

其中

$$A=\begin{pmatrix}0.67&0.67&0&0\\-0.67&0.67&0.67&0.67\\0.67&0.67&-0.67&-0.67\\0.67&0.67&0.67&-0.67\end{pmatrix}$$

$$B=\begin{pmatrix}0.6598\\1.9698\\4.3171\\-2.6436\end{pmatrix},\ K=\begin{pmatrix}-0.1027\\0.5501\\0.3545\\-0.5133\end{pmatrix}$$

$$C=(-0.5749,1.0751,-0.5225,0.1830),\ D=-0.7138$$

时间序列 $\boldsymbol{e}_k$ 是服从均值为 0, 方差为 9 的高斯白噪声, 即

$$e_k\sim N(0,9) \tag{22}$$

本节利用所提出的正交梯度二步递阶优化 (Hierarchical optimization method, HO method) 辨识算法分别进行了系统矩阵 A 的特征值分析、算法数值稳定性、算法效率以及模型一步预测性能的仿真实验.

4.1 系统参数矩阵 $\boldsymbol{A}$ 的特征值分析

噪声参数满足式 (22), 输入信号为服从 $\boldsymbol{u}_k\sim N(0,1)$ 的高斯白噪声序列, 且与噪声序列 $\boldsymbol{e}_k$ 统计独立, 每次实验中取不同的噪声序列而输入信号序列 $\boldsymbol{u}_k$ 相同. 分别用本文提出的递阶优化方法以及文献 [11] 的期望极大 (Expectation maximization, EM) 方法进行 30 次蒙特卡罗实验, 每次数值实验数据长度取为 1 000, 迭代步数为 35, 用 Matlab 内置函数 drss(·) 能够产生满足要求的能控能观单输入单输出四阶模型.

两种方法得到的矩阵 A 的实数特征值与迭代次数的关系如图 1 所示 (注: 矩阵 A 的虚数特征值与迭代次数的关系的结论类似, 证明从略).

从图 1 中可以得出结论: 本文提出的方法能以较快的学习速度收敛到系统矩阵 A 的两个实特征根真值 0.67 以及 −0.67.

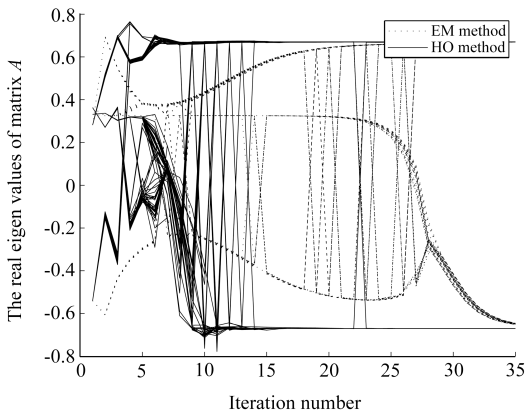

图 1 系统矩阵 A 的特征值与迭代次数关系

Fig. 1 The eigenvalues for matrix A with iterations

4.2 算法数值稳定性以及算法效率实验

对于本文提出的正交梯度二步递阶优化算法、Mckelvey 提出的基于数据驱动的参数局部调整辨识方法 (DDLC)[7] 以及 Bergboer[12], Verdult[13] 等学者给出的 LM 算法, 分别设计实验来分析三个算法的数值稳定性. 选取非线性最小二乘目标函数式 (6) 作为优化对象, 进行了 50 次蒙特卡罗实验. 输入信号是幅值为 1 的 PRBS (Pseudo-random binary sequence) 序列[19], 噪声参数满足式 (22), 每次仿真的数据长度为 1 000. 三种方法的目标函数最大值、最小值以及平均值 (分别用各自线型的上、中、下三根线表示) 与迭代步数的关系如图 2 所示. 从图 2 可以得出结论: 本文提出的方法收敛速度最快并且数值稳定性好.

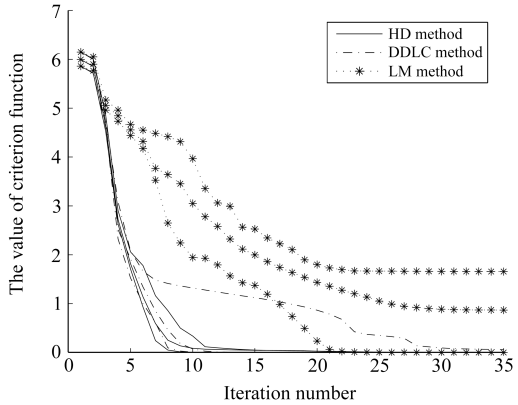

图 2 目标函数值与迭代次数的关系

Fig. 2 Values of the criterion function with iterations

为比较三个算法的效率, 分别记录了 50 次蒙特卡罗实验所消耗的机器时间. DDLC 方法以及正则化方法对本文提出的方法的时间比值关系如图 3 所示. 从图 3 可以得出如下结论: 每次数值实验中, DDLC 方法和 LM 方法对本文的正交梯度方法的机器时间比值均大于 1, 由此说明本文算法的效率最高.

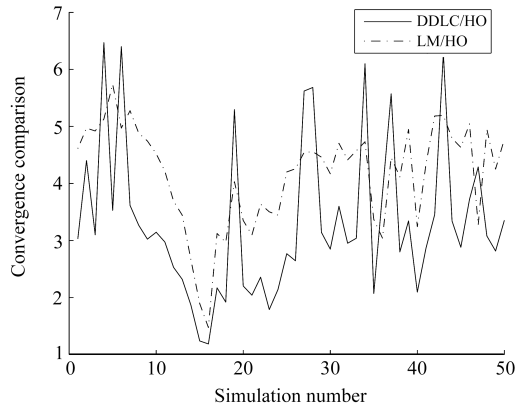

图 3 DDLC, LM 方法与本文方法的收敛速度比较

Fig. 3 Comparison of convergent speeds of the DDLC, LM, and HO methods

4.3 一步预测分析

为比较本文正交梯度递阶优化辨识方法和 Gibson 提出的期望极大 (EM) 方法的一步预测性能, 进行了 80 次蒙特卡罗实验. 输入信号是幅值为 1 的 PRBS 序列, 噪声参数满足式 (22), 两种方法的迭代次数均设为 35, 每次实验的数据长度为 2 000, 前 1 000 个数据用来系统建模, 后 1 000 个数据用来一步预测, 均方预测误差根据下式计算

$$MPE(\hat{\boldsymbol{\theta}}_k)=\frac{1}{lN}\sum_{p=1}^{N}\left(\boldsymbol{y}_p-\hat{\boldsymbol{y}}_{p|p-1}(\hat{\boldsymbol{\theta}}_k)\right)^{\mathrm{T}}\left(\boldsymbol{y}_p-\hat{\boldsymbol{y}}_{p|p-1}(\hat{\boldsymbol{\theta}}_k)\right)$$

其中, $l=1$ 是输出向量维数, $N=1\,000$ 是预测样本长度, $\hat{\boldsymbol{y}}_{p|p-1}(\hat{\boldsymbol{\theta}}_k)$ 是系统参数值为 $\hat{\boldsymbol{\theta}}_k$ 时系统输出的一步超前预测值. 每次实验中由子空间方法和期望极大 (EM) 方法得到的一步预测误差如图 4 (见下页) 所示.

图 4 中, 每个 "*" 代表每次蒙特卡罗实验的预测误差, 纵轴是本文方法的均方预测误差, 横轴是 EM 方法均预测误差, 位于对角线上方的 "*" 表示本文方法的预测误差大于 EM 方法的预测误差, 位于对角线下方的 "*" 表示本文方法的预测误差小于 EM 方法的预测误差.

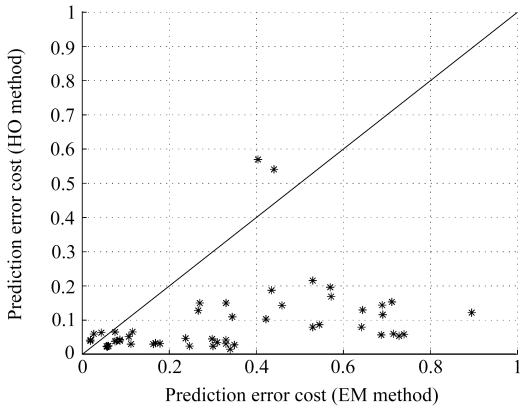

图 4　本文方法和 EM 方法的预测性能比较

Fig. 4　Comparison of prediction error costs for EM method and HO method

由图 4 可以看出, 只有 5 个 "*" 位于对角线上方, 这说明在 80 次实验中本文方法的预测误差小于 EM 方法的预测误差有 75 次, 表明本文方法得到的模型的预测精度高于 EM 方法, 本文方法估计得到的模型具有较好的一步预测能力.

5　结论

本文提出基于正交梯度的二步递阶优化方法来辨识一类测量噪声和过程噪声具有耦合关系的状态空间系统. 得到了沿随机状态空间系统输入 - 输出等价类相切平面的正交补空间优化系统参数的方法; 提出了由自适应 L-M 算法确定参数优化方向的方法, 并用一维搜索方法计算最佳步长. 仿真结果表明本文提出的正交梯度递阶优化辨识方法具有收敛速度快、抗噪能力强以及数值稳定性好等优点.

另外, 本文的辨识对象的噪声特点是: 输出噪声为高斯白噪声, 测量噪声和过程噪声满足线性耦合关系. 然而, 工业现场中系统所处的噪声通常具有强相关性以及非线性耦合, 为此, 后续的研究工作可以放在测量噪声和过程噪声具有更一般统计性质的状态系统辨识问题上.

References

1 Verhaegen M, Dewilde P. Subspace model identification part1: the output-error state-space model identification class of algorithms. *International Journal of Control*, 1992, **56**(5): 1187−1210

2 Van Overschee P, De Moor B. A unifying theorem for three subspace system identification algorithms. *Automatica*, 1995, **31**(12): 1853−1864

3 Qin S J, Lin W L, Ljung L. A novel subspace identification approach with enforced causal models. *Automatica*, 2005, **41**(12): 2043−2054

4 Favored W, De Moor B, Van Overschee P. Subspace state space system identification for industrial processes. *Journal of Process Control*, 2000, **10**(2): 149−155

5 Dahlén A, Lindquist A, Mari J. Experiment evidence showing that stochastic subspace identification methods may fail. *Systems and Control Letters*, 1998, **34**(5): 303−312

6 Ljung L. *System Identification: Theory for the User (Second Edition)*. New Jersey: Prentice-Hall, 1999

7 Mckelvey T, Helmersson A, Ribarits T. Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica*, 2004, **40**(9): 1629−1635

8 Ribarits T, Deistler M, Mckelvey T. An analysis of the parametrization by data driven local coordinates for multivariable linear systems. *Automatica*, 2004, **40**(5): 789−803

9 Ribarits T, Deistler M, Hanzon B. On new parametrization methods for the estimation of linear state-space models. *International Journal of Adaptive Control and Signal Processing*, 2004, **18**(9-10): 717−743

10 Ribarits T, Deistler M, Hanzon B. An analysis of separable least squares data driven local coordinates for maximum likelihood estimation of linear systems. *Automatica*, 2005, **41**(3): 531−544

11 Gibson S, Ninness B. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 2004, **41**(10): 1667−1682

12 Bergboer N H, Verdult V, Verhaegen M H G. An efficient implemention of maximum likelihood identification of LTI state-space models by local gradient search. In: Proceedings of the 41st IEEE Conference on Decision and Control. Las Vegas, USA: IEEE, 2002. 616−621

13 Verdult V, Bergboer N H, Verhaegen M H G. Maximum likelihood identification of multivariable bilinear state-space systems by projected gradient search. In: Proceedings of the 41st IEEE Conference on Decision and Control. Las Vegas, USA: IEEE, 2002. 1808−1813

14 Marquardt D W. An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 1963, **11**(2): 431−441

15 Lera G, Pinzolas M. Neighborhood based Levenberg-Marquardt algorithm for neural network training. *IEEE Transactions on Neural Networks*, 2002, **13**(5): 1200−1203

16 Gorinevsky D. An approach to parametric nonlinear least square optimization and application to task-level learning control. *IEEE Transactions on Automatic Control*, 1997, **42**(7): 912−927

17 Vanbleu K, Ysebaert G, Cuypers G, Moonen M. Adaptive bit rate maximizing time-domain equalizer design for DMT-based systems. *IEEE Transactions on Signal Processing*, 2006, **54**(2): 483−498

18 Dennis J E, Schnabel R B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New Jersey: Prentice-Hall, 1983

19 Xue Ding-Yu. *Simulation of Control System and Computer Aided Design*. Beijing: China Machine Press, 2005
(薛定宇. 控制系统仿真与计算机辅助设计. 北京: 机械工业出版社, 2005)

衷路生　浙江大学信息科学与工程学院博士研究生. 主要研究方向为线性系统辨识, 子空间辨识, 信号处理, 鲁棒估计. 本文通信作者.
E-mail: lszhongzju@sina.com
(**ZHONG Lu-Sheng**　Ph. D. candidate at College of Information Science and Engineering, Zhejiang University. His research interest covers linear stochastic system identification, subspace identification, signal processing, and robust estimation. Corresponding author of this paper.)

宋执环　浙江大学教授. 主要研究方向为信号处理, 故障诊断, 预测控制.
(**SONG Zhi-Huan**　Professor at Zhejiang University. His research interest covers signal processing, fault diagnosis, and predictive control.)