文章编号:1674-7070(2012)04-0289-12

辨识方法的计算效率(1):递推算法

丁锋[1,2,3]

摘要

算法的计算量可用其乘法运算次数和加法运算次数表示(除法作为乘法对待,减法作为加法对待).一次乘法运算或一次加法运算称为一个 flop,即一次浮点运算.作为"辨识方法的计算效率"系列3篇连载论文的第1篇,主要了讨论递推辨识算法的计算量,包括向量和矩阵基本运算的 flop 数,以及线性回归系统、多元线性回归系统、多变量系统的随机梯度辨识算法、最小二乘辨识算法、递推最小二乘辨识算法的最经济计算量,即实现算法的最少 flop 数.

关键词

递推辨识;迭代辨识;参数估计;FIR模型;方程误差模型;CAR 模型;CARMA 模型;CARAR 模型;CARARMA 模型;输出误差模型;OEMA 模型;OEAR 模型;辅助模型辨识;多新息辨识;递阶辨识;耦合辨识

中图分类号 TP273
文献标志码 A

收稿日期 2012-08-19
资助项目 国家自然科学基金(60973043);高等学校学科创新引智计划(B12018)

作者简介

丁锋,男,博士,教授,博士生导师,主要从事系统辨识、过程建模、自适应控制方面的研究. fding@ jiangnan. edu. cn

1 江南大学 物联网工程学院,无锡,214122
2 江南大学 控制科学与工程研究中心,无锡,214122
3 江南大学 教育部轻工过程先进控制重点实验室,无锡,214122

0 引言

数学模型是研究、分析、认识事物本质特征的工具,是求解优化问题、减小代价,进行设计、实行有效控制的基础. 系统辨识给系统建模提供了有效的方法和理论指导[1]. 数十年来,系统辨识得到了长足发展,从辨识的定义与辨识的基本步骤[2],系统描述的基本模型[3],辨识的基本问题[4],到辨识方法及其性能分析等,都进行了许多深入的研究,新的辨识方法不断问世,如辅助模型辨识方法[5]、迭代辨识方法[6]、多新息辨识方法[7]、递阶辨识方法[8]、耦合辨识方法[9]等. 因此,如何评价辨识算法的收敛性能,包括评价辨识算法的计算效率,估算辨识算法的计算量,分析参数估计的性质(收敛性、收敛速率)等都是辨识要解决的首要课题.

然而,随着科学技术的发展和电子设备计算能力的提高以及生产规模的扩大、控制和管理能力的增强,人们处理复杂、庞大对象的欲望越来越强烈. 对于控制系统而言,大系统或复杂系统的主要特征是输入输出变量多、强耦合、非线性、维数高、参数数目多,导致控制算法、参数辨识算法的复杂性和计算量增加,使算法的实现成为极为突出的问题. 这就需要分析辨识方法的计算效率,估算辨识算法的计算量,研制计算量小的辨识方法.

评价算法的计算量有很多方法. 例如,根据执行一个任务计算机运行时间的长短来评价. 因为问题的规模不同,使用的计算机也可能不同,不可能每次都用计算机来验证. 况且,有的算法需要计算机计算1年,有的需要计算2年,这就需要等上很长时间才能判断算法的计算量大小,显然这种方法不可取. 实际上,我们更期望不通过计算机运算就能估算和比较不同算法的计算量.

评价辨识算法计算量大小的一个实用方法是:用算法的乘法次数(除法作为乘法对待)和加法次数(减法作为加法对待)来衡量. 一次加法运算称为一个 flop,即一次浮点运算(float point operation),一次乘法运算也称为一个 flop[10]. 乘法运算次数和加法运算次数之和的 flop 数就是算法的计算量. 尽管数字大小的浮点运算不一样,但是用 flop 来刻画算法的计算量也不失为一种好方法. 如果同为递推算法(迭代算法),可以比较每一步递推计算(迭代计算)的 flop 数,来衡量算法的计算量.

值得注意的是,即使对于同一个算法,由于计算方式的不同,其计算量可能相差很大. 例如,多项式

$$f(x) = x^3 + 2x^2 + 3x + 4 = x \cdot x \cdot x + 2 \times x \cdot x + 3 \times x + 4$$

需要5次乘法运算和3次加法运算,如果将表达式修改为

$$f(x) = x(x(x+2)+3)+4,$$

就只需要3次乘法运算和3次加法运算. 因此,估算辨识方法的计算量,评价辨识方法的计算效率,需要找到能实现辨识算法的最经济计算方式,即实现辨识算法的最小 flop 数的计算方式,从而提高辨识算法的计算效率.

本文是"辨识方法的计算效率"系列3篇连载论文的第1篇——递推算法(另2篇是迭代算法和耦合算法),主要讨论向量和矩阵基本运算的 flop 数,以及线性回归系统、多元线性回归系统、多变量系统的随机梯度辨识算法、最小二乘辨识算法、递推最小二乘辨识算法实现的最经济计算量,即实现算法的最少 flop 数.

1 基本计算

首先定义一些符号. "$A=:X$"或"$X:=A$"表示"A 记作(定义为) X"之意;$\boldsymbol{I}$ 为适当维数的单位阵;$\boldsymbol{I}_n$ 为 n 阶单位阵;范数定义为 $\|\boldsymbol{X}\|^2:=\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$;$\mathbf{1}_{m\times n}$是其元均为1的 $m\times n$ 矩阵,即

$$\mathbf{1}_{m\times n}=\begin{bmatrix}1 & 1 & \cdots & 1\\ \vdots & \vdots & & \vdots\\ 1 & 1 & \cdots & 1\end{bmatrix}\in\mathbf{R}^{m\times n};$$

$\mathbf{1}_n:=\mathbf{1}_{n\times 1}$是元均为1的 n 维列向量;$\hat{\boldsymbol{\theta}}(t)$ 为参数向量 $\boldsymbol{\theta}$ 在时刻 t 的估计.

下面假设 α 和 β 是标量,即 $\alpha\in\mathbf{R}$,$\beta\in\mathbf{R}$,$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是实向量,$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是实矩阵.

1)标量数加和乘

标量数 $\alpha\in\mathbf{R}$ 与 $\beta\in\mathbf{R}$ 之和 $\alpha+\beta$ 有一次加法运算,其计算量为1 flop. 标量数 $\sigma\in\mathbf{R}$ 与 $\beta\in\mathbf{R}$ 之积 $\alpha\beta$ 有一次乘法运算,其计算量为1 flop.

2)标量与向量之积、标量与矩阵之积

标量 $\alpha\in\mathbf{R}$ 与向量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbf{R}^n$ 之积,即向量数乘(scalar-vector multiplication or vector scale)

$$\alpha\boldsymbol{x}=\alpha\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}=\begin{bmatrix}\alpha x_1\\ \alpha x_2\\ \vdots\\ \alpha x_n\end{bmatrix}\in\mathbf{R}^n$$

有 n 次乘法运算,因此其计算量为 n flops.

标量 $\beta\in\mathbf{R}$ 与矩阵 $\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$ 之积,即标量矩阵乘法(scalar-matrix multiplication)

$$\beta\boldsymbol{A}=\beta\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}=\begin{bmatrix}\beta a_{11} & \beta a_{12} & \cdots & \beta a_{1n}\\ \beta a_{21} & \beta a_{22} & \cdots & \beta a_{2n}\\ \vdots & \vdots & & \vdots\\ \beta a_{m1} & \beta a_{m2} & \cdots & \beta a_{mn}\end{bmatrix}\in\mathbf{R}^{m\times n}$$

有 mn 次乘法运算,其计算量为 mn flops.

表达式 $\alpha\boldsymbol{x}\beta=(\alpha\boldsymbol{x})\beta=\alpha(\boldsymbol{x}\beta)$ 有 $2n$ 次乘法运算,而 $\alpha\boldsymbol{x}\beta=(\alpha\beta)\boldsymbol{x}=\boldsymbol{x}(\alpha\beta)$ 只有 $n+1$ 次乘法运算.

表达式 $\alpha\boldsymbol{A}\beta=(\alpha\boldsymbol{A})\beta=\alpha(\boldsymbol{A}\beta)$ 有 $2mn$ 次乘法运算,而 $\alpha\boldsymbol{A}\beta=(\alpha\beta)\boldsymbol{A}=\boldsymbol{A}(\alpha\beta)$ 只有 $mn+1$ 次乘法运算.

由此可以看出,对于数学上等价的表达式,其先后计算方式不同,导致计算量有很大的差异.

3)向量之和、矩阵之和

向量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbf{R}^n$ 与 $\boldsymbol{y}=[y_1,y_2,\cdots,y_n]^{\mathrm{T}}\in\mathbf{R}^n$ 之和,即向量加法(vector addition)

$$\boldsymbol{x}+\boldsymbol{y}=\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}+\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}=\begin{bmatrix}x_1+y_1\\ x_2+y_2\\ \vdots\\ x_n+y_n\end{bmatrix}\in\mathbf{R}^n$$

有 n 次加法运算,即浮点运算次数(flop 数)为 n,所以计算量为 n flops.

矩阵 $\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$与矩阵 $\boldsymbol{B}=[b_{ij}]\in\mathbf{R}^{m\times n}$之和,即矩阵加法(matrix addition)

$$\boldsymbol{A}+\boldsymbol{B}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}+\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1n}\\ b_{21} & b_{22} & \cdots & b_{2n}\\ \vdots & \vdots & & \vdots\\ b_{m1} & b_{m2} & \cdots & b_{mn}\end{bmatrix}=\begin{bmatrix}a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n}\\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn}\end{bmatrix}\in\mathbf{R}^{m\times n}$$

有 mn 次加法运算,所以计算量为 mn flops.

表达式 $\alpha\boldsymbol{A}+\alpha\boldsymbol{B}$ 有 mn 次加法运算和 $2mn$ 次乘法运算,其计算量为 $3mn$ flops,但 $\alpha(\boldsymbol{A}+\boldsymbol{B})$ 有 mn 次加法运算和 mn 次乘法运算,其计算量为 $2mn$ flops,比 $\alpha\boldsymbol{A}+\alpha\boldsymbol{B}$ 的计算量小 mn flops.

由此可见,一个算法尽管其数学表达式是等价的,但由于计算方式的不同,其计算量也是不同的.因此,研究辨识算法的计算效率,需要研究辨识算法的计算方式,以及计算程序软件实现问题.

4) 向量点积或向量内积

向量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbf{R}^n$ 与 $\boldsymbol{y}=[y_1,y_2,\cdots,y_n]^{\mathrm{T}}\in\mathbf{R}^n$ 的点积(dot product)或内积(inner product)

$$\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=[x_1,x_2,\cdots,x_n]\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\sum_{i=1}^{n}x_iy_i\in\mathbf{R}$$

有 n 次乘法运算和 $n-1$ 次加法运算,flop 数分别为 n 和 $n-1$,其计算量为$(2n-1)$ flops. 表达式$(\alpha\boldsymbol{x}^{\mathrm{T}})\boldsymbol{y}$ 和 $\boldsymbol{x}^{\mathrm{T}}(\alpha\boldsymbol{y})$的计算量为$(3n-1)$ flops,$\alpha\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=\alpha(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})$ 的计算量为 $2n$ flops.

5) 向量外积

向量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_m]^{\mathrm{T}}\in\mathbf{R}^m$ 与 $\boldsymbol{y}=[y_1,y_2,\cdots,y_n]^{\mathrm{T}}\in\mathbf{R}^n$ 的外积(outer prodcut)

$$\boldsymbol{x}\times\boldsymbol{y}=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_m\end{bmatrix}[y_1,y_2,\cdots,y_n]=$$

$$\begin{bmatrix}x_1y_1 & x_1y_2 & \cdots & x_1y_n\\x_2y_1 & x_2y_2 & \cdots & x_2y_n\\\vdots & \vdots & & \vdots\\x_my_1 & x_my_2 & \cdots & x_my_n\end{bmatrix}\in\mathbf{R}^{m\times n}$$

有 mn 次乘法运算,计算量为 mn flops. 表达式 $\alpha\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}=\alpha(\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}})$的计算量为 $2mn$ flops,$(\alpha\boldsymbol{x})\boldsymbol{y}^{\mathrm{T}}$ 的计算量为$(mn+m)$ flops,$\boldsymbol{x}(\boldsymbol{y}^{\mathrm{T}}\alpha)$的计算量为$(mn+n)$ flops.

6) 矩阵与向量之积

设 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{y}=[y_1,y_2,\cdots,y_m]^{\mathrm{T}}\in\mathbf{R}^m$,$\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$. 矩阵与向量乘法(matrix-vector multiplication),即表达式

$$\boldsymbol{A}\boldsymbol{x}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & & \vdots\\a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}=$$

$$\begin{bmatrix}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\\vdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\end{bmatrix}\in\mathbf{R}^m$$

有 mn 次乘法运算和 $m(n-1)$次加法运算,其计算量为$(2mn-m)$ flops. $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{y}$ 的计算量为 $2mn$ flops. 表达式$(\alpha\boldsymbol{A})\boldsymbol{x}$ 的计算量为$(3mn-m)$ flops,但 $\boldsymbol{A}(\boldsymbol{x}\alpha)$的计算量为$(2mn-m+n)$ flops. 表达式 $\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x})$有 $mn+m$ 次乘法运算和 $m(n-1)+m-1=mn-1$ 次加法运算,其计算量为$(2mn+m-1)$ flops,$(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x}$ 的计算量为$(2mn+n-1)$ flops.

7) 矩阵与矩阵之积

设 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbf{R}^n$,$\boldsymbol{y}=[y_1,y_2,\cdots,y_m]^{\mathrm{T}}\in\mathbf{R}^m$,$\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$,$\boldsymbol{B}=[b_{ij}]\in\mathbf{R}^{n\times p}$. 矩阵与矩阵乘法(matrix-matrix multiplication),即矩阵乘积表达式

$$\boldsymbol{A}\boldsymbol{B}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & & \vdots\\a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1p}\\b_{21} & b_{22} & \cdots & b_{2p}\\\vdots & \vdots & & \vdots\\b_{n1} & b_{n2} & \cdots & b_{np}\end{bmatrix}=$$

$$\begin{bmatrix}a_{11}b_{11}+\cdots+a_{1n}b_{n1} & \cdots & a_{11}b_{1p}+\cdots+a_{1n}b_{np}\\a_{21}b_{11}+\cdots+a_{2n}b_{n1} & \cdots & a_{21}b_{1p}+\cdots+a_{2n}b_{np}\\\vdots & & \vdots\\a_{m1}b_{11}+\cdots+a_{mn}b_{n1} & \cdots & a_{m1}b_{1p}+\cdots+a_{mn}b_{np}\end{bmatrix}\in\mathbf{R}^{m\times p}$$

有 mnp 次乘法运算和 $m(n-1)p$ 次加法运算,其计算量为$(2mnp-mp)$ flops. 也就是说,矩阵乘法的计算量数量级为 $O(mnp)$.

对于同维数矩阵 $\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$和 $\boldsymbol{B}=[b_{ij}]\in\mathbf{R}^{m\times n}$,$\boldsymbol{A}\boldsymbol{B}$ 和 $\boldsymbol{A}^2$ 均有 n^3 次乘法运算和 n^3-n^2 次加法运算,其计算量为$(2n^3-n^2)$ flops,计算量数量级为 $O(n^3)$. 尽管矩阵乘法计算大,但是矩阵的行列式、矩阵的特征值、矩阵逆的计算量更大.

下面引入一个新的运算符——冒号运算符(colon notation). 它能极为方便地表示矩阵的一列或一行和子矩阵.

设 $\boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}$,$(\boldsymbol{A})_{ij}$表示 $\boldsymbol{A}$ 的第(i,j)元,即 $(\boldsymbol{A})_{ij}=a_{ij}$. $\boldsymbol{A}(i,:)$表示 $\boldsymbol{A}$ 的第 i 行(这里冒号可以理解为"所有"或"任意"),即

$$\boldsymbol{A}(i,:):=[a_{i1},a_{i2},\cdots,a_{in}]\in\mathbf{R}^{1\times n}.$$

$\boldsymbol{A}(:,j)$表示 $\boldsymbol{A}$ 的第 j 列,即

$$\boldsymbol{A}(:,j):=\begin{bmatrix}a_{1j}\\a_{2j}\\\vdots\\a_{mj}\end{bmatrix}\in\mathbf{R}^m.$$

用 $\boldsymbol{A}(i:j,p:q)$表示表示从第 i 到第 j 行,第 p 到 q 列的子矩阵(submatrix)(这里冒号可以理解为"到"),即

$\boldsymbol{A}(i:j,p:q):=$

$$\begin{bmatrix} a_{i,p} & a_{i,p+1} & a_{i,p+2} & \cdots & a_{i,q} \\ a_{i+1,p} & a_{i+1,p+1} & a_{i+1,p+2} & \cdots & a_{i+1,q} \\ a_{i+2,p} & a_{i+2,p+1} & a_{i+2,p+2} & \cdots & a_{i+2,q} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{j,p} & a_{j,p+1} & a_{j,p+2} & \cdots & a_{j,q} \end{bmatrix} \in \mathbf{R}^{(j-i+1)\times(q-p+1)}.$$

类似地,有

$\boldsymbol{A}(:,p:q):=$

$$\begin{bmatrix} a_{1,p} & a_{1,p+1} & a_{1,p+2} & \cdots & a_{1,q} \\ a_{2,p} & a_{2,p+1} & a_{2,p+2} & \cdots & a_{2,q} \\ a_{3,p} & a_{3,p+1} & a_{3,p+2} & \cdots & a_{3,q} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m,p} & a_{m,p+1} & a_{m,p+2} & \cdots & a_{m,q} \end{bmatrix} \in \mathbf{R}^{m\times(q-p+1)},$$

$\boldsymbol{A}(i:j,:):=$

$$\begin{bmatrix} a_{i,1} & a_{i,2} & a_{i,3} & \cdots & a_{i,n} \\ a_{i+1,1} & a_{i+1,2} & a_{i+1,3} & \cdots & a_{i+1,n} \\ a_{i+2,1} & a_{i+2,2} & a_{i+2,3} & \cdots & a_{i+2,n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{j,1} & a_{j,2} & a_{j,3} & \cdots & a_{j,n} \end{bmatrix} \in \mathbf{R}^{(j-i+1)\times n}.$$

例如

$$\boldsymbol{B}:=\boldsymbol{A}(2:4,3:6)=\begin{bmatrix} a_{23} & a_{24} & a_{25} & a_{26} \\ a_{33} & a_{34} & a_{35} & a_{36} \\ a_{43} & a_{44} & a_{45} & a_{46} \end{bmatrix} \in \mathbf{R}^{3\times 4},$$

那么

$$\boldsymbol{B}(:,1:2)==\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}=\begin{bmatrix} a_{23} & a_{24} \\ a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix} \in \mathbf{R}^{3\times 2}.$$

2 线性回归系统(LR)

2.1 随机梯度算法(SG)

考虑线性回归模型(linear regression model)

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t), \tag{1}$$

其中$\{y(t)\}$是观测输出序列,$\{v(t)\}$是零均值不相关随机噪声序列,$\boldsymbol{\varphi}(t)\in\mathbf{R}^n$是观测数据构成的可测信息向量,$\boldsymbol{\theta}\in\mathbf{R}^n$是要估计的参数向量. 假设$t\leqslant 0$时,$y(t)=0$,$\boldsymbol{\varphi}(t)=\mathbf{0}$和$v(t)=0$.

估计系统(1)中参数向量$\boldsymbol{\theta}$的随机梯度算法(Stochastic Gradient algorithm, SG算法)如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0>0, \tag{2}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1. \tag{3}$$

式(2)有下列2种计算方式(式中多余的括号表示计算顺序):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\left\{\frac{\boldsymbol{\varphi}(t)}{r(t)}\right\}e(t), \tag{4}$$

$$e(t):=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{5}$$

和

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{\varphi}(t)\left\{\frac{e(t)}{r(t)}\right\}. \tag{6}$$

对于递推算法(迭代算法),只需列出每一步递推计算(迭代计算)的计算量. 式(4)的乘法次数为$2n$,加法次数为n;式(6)的乘法次数为$n+1$,加法次数为n. 故式(6)比式(4)的计算量小. 因此,存在一种计算方式,使得算法的计算量达到最小. 式(5)新息$e(t)$的乘法次数为n,加法次数为n;式(3)的乘法次数为n,加法次数也为n. 因此,算法(3)—(5)的乘法次数为$4n$,加法次数也为$3n$,计算量为$7n$ flops;算法(3)和(5)—(6)的乘法次数为$3n+1$,加法次数也为$3n$,计算量为$(6n+1)$ flops. 因此,随机梯度算法的最小计算量为$(6n+1)$ flops. 表1列出了SG算法(2)—(3)每一步递推计算中的乘法次数、加法次数和flop数.

2.2 最小二乘算法(LS)

考虑式(1)的线性回归模型(linear regression model),重写如下:

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}+v(t). \tag{7}$$

定义二次准则函数(quadratic criterion function)

表1 随机梯度算法的计算量

Table 1 The computational efficiency of the SG algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$e(t):=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}$	n	n
	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{\varphi}(t)[e(t)/r(t)]\in\mathbf{R}^n$	$n+1$	n
$r(t)$	$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2\in\mathbf{R}$	n	n
总数		$3n+1$	$3n$
总flop数		$6n+1$	

$$J_1(\boldsymbol{\theta}):=\sum_{j=1}^{t}[y(j)-\boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2.$$

参考文献[4],令 $J_1(\boldsymbol{\theta})$ 对 $\boldsymbol{\theta}$ 的偏导数为零,得到参数向量 $\boldsymbol{\theta}$ 的最小二乘估计(Least Squares Estimate, LSE 估计):

$$\hat{\boldsymbol{\theta}}(t)=\left[\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\right]^{-1}\left[\sum_{j=1}^{t}\boldsymbol{\varphi}(j)y(j)\right].\quad(8)$$

这是最小二乘估计一次完成算法(direct algorithm).对于每一个 t,式(8)都要计算矩阵逆,计算量很大.下面用递推方式来实现(计算)这个最小二乘估计.值得指出的是,与递推最小二乘算法的不同,这里的递推不是参数估计计算式的递推,而是一些中间变量的递推计算.

定义协方差阵 $\boldsymbol{P}(t)$ 和向量 $\boldsymbol{\xi}(t)$ 如下:

$$\boldsymbol{P}^{-1}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j),$$
$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n>0,\quad(9)$$

$$\boldsymbol{\xi}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)y(j)=\boldsymbol{\xi}(t-1)+\boldsymbol{\varphi}(t)y(t),$$
$$\boldsymbol{\xi}(0)=\boldsymbol{0}.\quad(10)$$

式(8)的最小二乘估计可以表示为

$$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\xi}(t).\quad(11)$$

将矩阵求逆引理(matrix inversion lemma)

$$(\boldsymbol{A}+\boldsymbol{BC})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{CA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{CA}^{-1}\quad(12)$$

应用到式(9),可得

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}.\quad(13)$$

式(11),(13)和(10)构成了最小二乘算法(Least Squares algorithm, LS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\xi}(t),\quad(14)$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)},$$
$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n,\quad(15)$$

$$\boldsymbol{\xi}(t)=\boldsymbol{\xi}(t-1)+\boldsymbol{\varphi}(t)y(t),\quad\boldsymbol{\xi}(0)=\boldsymbol{0}.\quad(16)$$

令 $\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$. 注意到 $\boldsymbol{P}(t)$ 是对称阵(symmetric matrix),上式可以表示为

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{\zeta}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)}=$$
$$\boldsymbol{P}(t-1)-\left\{\frac{\boldsymbol{\zeta}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)}\right\}\boldsymbol{\zeta}^{\mathrm{T}}(t).\quad(17)$$

表2为最小二乘算法的计算量.

2.3 递推最小二乘算法(RLS)

参考文献[4],式(7)参数向量 $\boldsymbol{\theta}$ 的递推最小二乘算法(Recursive Least Squares algorithm, RLS 算法)为

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad(18)$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),$$
$$\boldsymbol{P}(0)=p_0\boldsymbol{I}>0.\quad(19)$$

定义增益向量 $\boldsymbol{L}(t):=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$. 式(13)两边右乘向量 $\boldsymbol{\varphi}(t)$ 可得

$$\boldsymbol{L}(t)=\boldsymbol{P}(t)\boldsymbol{\varphi}(t)=\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}=$$
$$\frac{\boldsymbol{\zeta}(t)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)}.\quad(20)$$

借助于增益向量 $\boldsymbol{L}(t)$,式(13)中 $\boldsymbol{P}(t)$ 有下列几种形式:

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]\boldsymbol{L}^{\mathrm{T}}(t)$$
$$=[\boldsymbol{I}-\boldsymbol{L}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{P}(t-1)\quad(21)$$
$$=\boldsymbol{P}(t-1)[\boldsymbol{I}-\boldsymbol{\varphi}(t)\boldsymbol{L}^{\mathrm{T}}(t)]\quad(22)$$
$$=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)[\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]^{\mathrm{T}}.\quad(23)$$

式(21)或(22)的计算量为$(2n^3+n)$ flops,其中乘法运算次数为 n^3+n^2 和加法运算次数为 n^3-n^2+n. 计算量是 n 的立方,即 $O(n^3)$,这种计算方法计算量太大,不可取.计算量最小的 RLS 算法表达如下:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)],$$
$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\quad(24)$$

表2 最小二乘算法的计算量

Table 2 The computational efficiency of the least squares algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\xi}(t)\in\mathbf{R}^n$	n^2	n^2-n
	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$	n^2	n^2-n
$\boldsymbol{P}(t)$	$\boldsymbol{L}(t):=\boldsymbol{\zeta}(t)/[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^n$	$2n$	n
	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	n^2	n^2
$\boldsymbol{\xi}(t)$	$\boldsymbol{\xi}(t)=\boldsymbol{\xi}(t-1)+\boldsymbol{\varphi}(t)y(t)\in\mathbf{R}^n$	n	n
	总数	$3n^2+3n$	$3n^2$
	总 flop 数	$6n^2+3n$	

$$\boldsymbol{L}(t)=\frac{[\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)[\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]},\tag{25}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)[\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]^{\mathrm{T}},\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n.\tag{26}$$

RLS 算法(24)—(26)的计算量为$(4n^2+6n)$ flops,如表 3 所示[4]. RLS 算法的计算量比表 2 中最小二乘估计的计算量小$(6n^2+3n)-(4n^2+6n)=2n^2-3n$ 次,所以递推最小二乘算法的计算量比一次完成最小二乘算法计算量小.

3 多元线性回归系统(MLR)

考虑下列多元线性回归系统(multivariate linear regression system)

$$\boldsymbol{y}(t)=\boldsymbol{\Phi}(t)\boldsymbol{\theta}+\boldsymbol{v}(t),\tag{27}$$

其中$\boldsymbol{y}(t)=[y_1(t),y_2(t),\cdots,y_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 为 m 维系统输出向量,$\boldsymbol{\Phi}(t)\in\mathbf{R}^{m\times n}$是由系统输入输出数据构成的回归信息矩阵(通常 $n>m$),$\boldsymbol{\theta}\in\mathbf{R}^n$ 是待辨识的系统参数向量,$\boldsymbol{v}(t)=[v_1(t),v_2(t),\cdots,v_m(t)]^{\mathrm{T}}\in\mathbf{R}^m$ 是零均值白噪声向量. 假设 $t\leqslant 0$ 时,$\boldsymbol{y}(t)=\mathbf{0}$,$\boldsymbol{\Phi}(t)=0$ 和 $\boldsymbol{v}(t)=0$.

3.1 多元随机梯度算法(MSG)

定义和极小化梯度准则函数(gradient criterion function)

$$J_2(\boldsymbol{\theta}):=\|\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\boldsymbol{\theta}\|^2,$$

可得估计参数向量$\boldsymbol{\theta}$的多元随机梯度算法(Multivariate Stochastic Gradient algorithm,MSG 算法)[9,11]:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Phi}^{\mathrm{T}}(t)}{R(t)}[\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_n/p_0,\tag{28}$$

$$R(t)=R(t-1)+\|\boldsymbol{\Phi}(t)\|^2,\quad R(0)=1.\tag{29}$$

多元随机梯度估计算法(28)—(29)的计算量为$(6mn+m)$ flops,如表 4 所示.

3.2 多元最小二乘算法(MLS)

定义和极小化最小二乘准则函数(least squares criterion function)

$$J_3(\boldsymbol{\theta}):=\sum_{j=1}^{t}\|\boldsymbol{y}(j)-\boldsymbol{\Phi}(j)\boldsymbol{\theta}\|^2.$$

令 $J_3(\boldsymbol{\theta})$对$\boldsymbol{\theta}$的偏导数为零,得到参数向量$\boldsymbol{\theta}$的多元最小二乘估计:

$$\hat{\boldsymbol{\theta}}(t)=\left[\sum_{j=1}^{t}\boldsymbol{\Phi}^{\mathrm{T}}(j)\boldsymbol{\Phi}(j)\right]^{-1}\left[\sum_{j=1}^{t}\boldsymbol{\Phi}^{\mathrm{T}}(j)\boldsymbol{y}(j)\right].\tag{30}$$

定义协方差阵$\boldsymbol{P}(t)$和向量$\boldsymbol{\xi}(t)$如下:

$$\boldsymbol{P}^{-1}(t):=\sum_{j=1}^{t}\boldsymbol{\Phi}^{\mathrm{T}}(j)\boldsymbol{\Phi}(j)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(j)\boldsymbol{\Phi}(j)\in\mathbf{R}^{n\times n},\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n,\tag{31}$$

$$\boldsymbol{\xi}(t):=\sum_{j=1}^{t}\boldsymbol{\Phi}^{\mathrm{T}}(j)\boldsymbol{y}(j)=\boldsymbol{\xi}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{y}(t)\in\mathbf{R}^n,\quad \boldsymbol{\xi}(0)=\mathbf{0}.\tag{32}$$

计算式(30)最小二乘估计$\hat{\boldsymbol{\theta}}(t)$的多元最小二乘算法

表 3 RLS 算法每步的计算量

Table 3 The computational efficiency of the RLS algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$e(t):=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}$	n	n
	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)e(t)\in\mathbf{R}^n$	n	n
$\boldsymbol{L}(t)$	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$	n^2	$(n-1)n$
	$\boldsymbol{L}(t)=\boldsymbol{\zeta}(t)/[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^n$	$2n$	n
$\boldsymbol{P}(t)$	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	n^2	n^2
	总　数	$2n^2+4n$	$2n^2+2n$
	总 flop 数	$4n^2+6n$	

表 4 多元随机梯度算法的计算量

Table 4 The computational efficiency of the multivariate stochastic gradient algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\boldsymbol{e}(t):=\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}^m$	mn	mn
	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{e}(t)R(t)]\in\mathbf{R}^n$	$mn+m$	mn
$R(t)$	$R(t)=R(t-1)+\|\boldsymbol{\Phi}(t)\|^2\in\mathbf{R}$	mn	mn
	总　数	$3mn+m$	$3mn$
	总 flop 数	$6mn+m$	

(Multivariate Least Squares algorithm，MLS 算法)如下：

$$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\xi}(t), \tag{33}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{I}_m+\boldsymbol{\Phi}(t)\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)]^{-1}\boldsymbol{\Phi}(t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{34}$$

$$\boldsymbol{\xi}(t)=\boldsymbol{\xi}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{y}(t),\quad \boldsymbol{\xi}(0)=\boldsymbol{0}. \tag{35}$$

多元最小二乘估计算法(33)—(35)的计算量为$[(4m+2)n^2+4m^2n-m^2-n+m+x+y]$ flops，如表5所示，其中x和y表示对矩阵$\boldsymbol{\Lambda}(t)$求逆的乘法运算次数与加法运算次数.

3.3 多元递推最小二乘算法(MRLS)

参考文献[4,9]，可以得到计算式(30)参数估计$\hat{\boldsymbol{\theta}}(t)$的递推最小二乘算法：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\boldsymbol{1}_n/p_0, \tag{36}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\Phi}(t),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n. \tag{37}$$

为避免计算式(37)大矩阵$\boldsymbol{P}(t)\in\mathbf{R}^{n\times n}$的逆(一般$n>m$)，应用矩阵求逆引理于式(37)，能够得到等价的估计多元线性回归系统(27)参数向量$\boldsymbol{\theta}$的多元递推最小二乘算法(Multivariate Recursive Least Squares algorithm，MRLS 算法)：

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)],\quad \hat{\boldsymbol{\theta}}(0)=\boldsymbol{1}_n/p_0, \tag{38}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{I}_m+\boldsymbol{\Phi}(t)\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)]^{-1}, \tag{39}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\Phi}(t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n. \tag{40}$$

多元递推最小二乘估计算法(38)—(40)的计算量为$(4mn^2+4m^2n+2mn-m^2+m+x+y)$ flops，如表6所示. 多元递推最小二乘算法的计算量比多元最小二乘估计的计算量小$[(4m+2)n^2+4m^2n-m^2-n+m+x+y]-[4mn^2+4m^2n+2mn-m^2+m+x+y]=(2n^2-2mn-n)$ flops.

表5 多元最小二乘估计算法的计算量

Table 5 The computational efficiency of the multivariate least squares algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\xi}(t)\in\mathbf{R}^n$	n^2	n^2-n
	$\boldsymbol{Q}(t):=\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)\in\mathbf{R}^{m\times m}$	mn^2	mn^2-mn
	$\boldsymbol{\Lambda}(t):=\boldsymbol{I}\_m+\boldsymbol{\Phi}(t)\boldsymbol{Q}(t)\in\mathbf{R}^{m\times m}$	m^2n	m^2n-m^2+m
$\boldsymbol{P}(t)$	$\boldsymbol{\Lambda}'(t):=\boldsymbol{\Lambda}^{-1}(t)\in\mathbf{R}^{m\times m}$	x	y
	$\boldsymbol{L}(t):=\boldsymbol{Q}(t)\boldsymbol{\Lambda}'(t)\in\mathbf{R}^{n\times m}$	m^2n	m^2n-mn
	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{Q}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	mn^2	mn^2
$\boldsymbol{\xi}(t)$	$\boldsymbol{\xi}(t)=\boldsymbol{\xi}(t-1)+\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{y}(t)\in\mathbf{R}^n$	mn	mn
	总　数	$(2m+1)n^2+2m^2n+mn+x$	$(2m+1)n^2+2m^2n-mn-m^2-n+m+y$
	总 flop 数	$(4m+2)n^2+4m^2n-m^2-n+m+x+y$	

表6 多元递推最小二乘算法的计算量

Table 6 The computational efficiency of the multivariate recursive least squares algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\boldsymbol{e}(t):=\boldsymbol{y}(t)-\boldsymbol{\Phi}(t)\hat{\boldsymbol{\theta}}(t-1)\in\mathbf{R}^m$	mn	mn
	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)\boldsymbol{e}(t)\in\mathbf{R}^n$	mn	mn
	$\boldsymbol{Q}(t):=\boldsymbol{P}(t-1)\boldsymbol{\Phi}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times m}$	mn^2	mn^2-mn
$\boldsymbol{L}(t)$	$\boldsymbol{\Lambda}(t):=\boldsymbol{I}_m+\boldsymbol{\Phi}(t)\boldsymbol{Q}(t)\in\mathbf{R}^{m\times m}$	m^2n	m^2n-m^2+m
	$\boldsymbol{\Lambda}'(t):=\boldsymbol{\Lambda}^{-1}(t)\in\mathbf{R}^{m\times m}$	x	y
	$\boldsymbol{L}(t):=\boldsymbol{Q}(t)\boldsymbol{\Lambda}'(t)\in\mathbf{R}^{n\times m}$	m^2n	m^2n-mn
$\boldsymbol{P}(t)$	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{Q}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	mn^2	mn^2
	总　数	$2mn^2+2m^2n+2mn+x$	$2mn^2+2m^2n-m^2+m+y$
	总 flop 数	$4mn^2+4m^2n+2mn-m^2+m+x+y$	

4 多输入多输出系统(MIMO)

考虑下列多输入多输出系统(Multi-Input Multi-Output system, MIMO 系统),即多变量系统(multivariable system)[12-13]

$$\boldsymbol{A}(z)\boldsymbol{y}(t)=\boldsymbol{B}(z)\boldsymbol{u}(t)+\boldsymbol{v}(t),$$

其中 $\boldsymbol{u}(t)\in\mathbf{R}^r$ 为系统输入向量,$\boldsymbol{y}(t)\in\mathbf{R}^m$ 为系统输出向量,$\boldsymbol{v}(t)\in\mathbf{R}^m$ 为零均值随机噪声向量,$\boldsymbol{A}(z)$ 和 $\boldsymbol{B}(z)$ 均为单位后移算子 z^{-1} 的多项式矩阵 $[z^{-1}\boldsymbol{y}(t)=\boldsymbol{y}(t-1)]$,且

$$\boldsymbol{A}(z):=\boldsymbol{I}+\boldsymbol{A}_1z^{-1}+\boldsymbol{A}_2z^{-2}+\cdots+\boldsymbol{A}_{n_a}z^{-n_a},$$

$$\boldsymbol{B}(z):=\boldsymbol{B}_1z^{-1}+\boldsymbol{B}_2z^{-2}+\cdots+\boldsymbol{B}_{n_b}z^{-n_b}.$$

假设阶次 n_a 和 n_b 已知,$\{\boldsymbol{u}(t),\boldsymbol{y}(t)\}$ 是观测输入输出数据,$\boldsymbol{A}_i\in\mathbf{R}^{m\times m}$ 和 $\boldsymbol{B}_i\in\mathbf{R}^{m\times r}$ 是待辨识的系统参数矩阵.

定义参数矩阵 $\boldsymbol{\theta}$ 和信息向量 $\boldsymbol{\varphi}(t)$ 如下:

$$\boldsymbol{\theta}^{\mathrm{T}}:=[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_{n_a},\boldsymbol{B}_1,\cdots,\boldsymbol{B}_{n_b}]\in\mathbf{R}^{m\times n},$$

$$n:=mn_a+rn_b,$$

$$\boldsymbol{\varphi}(t):=[-\boldsymbol{y}^{\mathrm{T}}(t-1),-\boldsymbol{y}^{\mathrm{T}}(t-2),\cdots,-\boldsymbol{y}^{\mathrm{T}}(t-n_a),\boldsymbol{u}^{\mathrm{T}}(t-1),\boldsymbol{u}^{\mathrm{T}}(t-2),\cdots,\boldsymbol{u}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}\in\mathbf{R}^n.$$

则式(41)可等价写为

$$\boldsymbol{y}(t)=\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t)+\boldsymbol{v}(t). \tag{42}$$

式(42)称为多变量系统的辨识模型或辨识表达式.

4.1 多变量随机梯度算法(M-SG)

定义和极小化梯度准则函数(gradient criterion function)

$$J_4(\boldsymbol{\theta}):=\|\boldsymbol{y}(t)-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t)\|^2,$$

可得估计系统(42)参数矩阵 $\boldsymbol{\theta}$ 的多变量随机梯度算法(Multivariable Stochastic Gradient algorithm, M-SG)算法:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\varphi}(t)}{r(t)}\boldsymbol{e}^{\mathrm{T}}(t),$$

$$\hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n\times m}/p_0, \tag{43}$$

$$\boldsymbol{e}(t)=\boldsymbol{y}(t)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t), \tag{44}$$

$$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2,\quad r(0)=1. \tag{45}$$

多变量随机梯度辨识算法(43)—(45)的计算量为 $(4mn+m+2n)$ flops,如表 7 所示.

4.2 多变量最小二乘算法(M-LS)

令 $J_4(\boldsymbol{\theta})$ 对 $\boldsymbol{\theta}$ 的偏导数为零,得到参数向量 $\boldsymbol{\theta}$ 的多变量最小二乘估计:

$$\hat{\boldsymbol{\theta}}(t)=\left[\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\right]^{-1}\left[\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{y}^{\mathrm{T}}(j)\right]. \tag{46}$$

定义协方差阵 $\boldsymbol{P}(t)$ 和矩阵 $\boldsymbol{\Xi}(t)$ 如下:

$$\boldsymbol{P}^{-1}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(j)\boldsymbol{\varphi}^{\mathrm{T}}(j)\in\mathbf{R}^{n\times n},\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{47}$$

$$\boldsymbol{\Xi}(t):=\sum_{j=1}^{t}\boldsymbol{\varphi}(j)\boldsymbol{y}^{\mathrm{T}}(j)=\boldsymbol{\Xi}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{y}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times m},\quad \boldsymbol{\Xi}(0)=\mathbf{0}. \tag{48}$$

计算式(46)最小二乘估计 $\hat{\boldsymbol{\theta}}(t)$ 的多变量最小二乘算法(Multivariable Least Squares algorithm, M-LS 算法)如下:

$$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\Xi}(t), \tag{49}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)},$$

$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n, \tag{50}$$

$$\boldsymbol{\Xi}(t)=\boldsymbol{\Xi}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{y}^{\mathrm{T}}(t),\quad \boldsymbol{\Xi}(0)=\mathbf{0}. \tag{51}$$

多变量最小二乘辨识算法(49)—(51)的计算量为 $(m+2)n(2n+1)$ flops,如表 8 所示.

4.3 多变量递推最小二乘算法(M-RLS)

参考文献[4,9],可以得到计算式(46)参数估计 $\hat{\boldsymbol{\theta}}(t)$ 的递推最小二乘算法:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}(t)\boldsymbol{\varphi}(t)[\boldsymbol{y}(t)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t)]^{\mathrm{T}}, \tag{52}$$

$$\boldsymbol{P}^{-1}(t)=\boldsymbol{P}^{-1}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t),$$

$$\boldsymbol{P}(0)=p_0\boldsymbol{I}_n. \tag{53}$$

为避免计算式(53)矩阵 $\boldsymbol{P}(t)\in\mathbf{R}^{n\times n}$ 的逆,应用矩阵求逆引理于式(53),能够得到等价的估计多变量线性回归系统(42)参数向量 $\boldsymbol{\theta}$ 的多变量递推最小

表 7 多变量随机梯度算法的计算量

Table 7 The computational efficiency of the multivariable stochastic gradient algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{\varphi}(t)[\boldsymbol{e}^{\mathrm{T}}(t)/r(t)]\in\mathbf{R}^{n\times m}$	$mn+m$	mn
$\boldsymbol{e}(t)$	$\boldsymbol{e}(t)=\boldsymbol{y}(t)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^m$	mn	mn
$r(t)$	$r(t)=r(t-1)+\|\boldsymbol{\varphi}(t)\|^2\in\mathbf{R}$	n	n
总 数		$2mn+m+n$	$2mn+n$
总 flop 数		$4mn+m+2n$	

二乘算法(Multivariable Recursive Least Squares algorithm,M-RLS 算法):

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)[\boldsymbol{y}(t)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t)]^{\mathrm{T}},\quad \hat{\boldsymbol{\theta}}(0)=\mathbf{1}_{n\times m}/p_0, \tag{54}$$

$$\boldsymbol{L}(t)=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)]^{-1}, \tag{55}$$

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{P}(t-1),\quad \boldsymbol{P}(0)=p_0\boldsymbol{I}_n. \tag{56}$$

多变量递推最小二乘辨识算法(54)—(56)的计算量为$(4n^2+4mn+2n)$ flops,如表 9 所示. 多变量递推最小二乘算法的计算量比多变量最小二乘估计的计算量小$[(m+2)n(2n+1)]-[4n^2+4mn+2n]=(2n-3)mn$ flops.

5 结语

与辨识算法的收敛速率、参数估计精度一样,辨识算法的计算效率与计算量也是评价辨识算法性能的重要指标. 本文主要讨论了线性回归系统、多元线性回归系统、多变量系统的随机梯度辨识算法、最小二乘辨识算法、递推最小二乘辨识算法的计算量. 进一步可以研究辅助模型辨识算法[14-20]、多新息辨识方法[21-35]、递阶辨识方法[36-50]、耦合辨识方法[9,51]等的计算效率.

表 8 多变量最小二乘估计算法的计算量

Table 8 The computational efficiency of the multivariable least squares algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\hat{\boldsymbol{\theta}}(t)=\boldsymbol{P}(t)\boldsymbol{\Xi}(t)\in\mathbf{R}^{n\times m}$	mn^2	mn^2-mn
$\boldsymbol{P}(t)$	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$	n^2	n^2-n
	$\boldsymbol{L}(t):=\boldsymbol{\zeta}(t)/[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^n$	$2n$	n
	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	n^2	n^2
$\boldsymbol{\Xi}(t)$	$\boldsymbol{\Xi}(t)=\boldsymbol{\Xi}(t-1)+\boldsymbol{\varphi}(t)\boldsymbol{y}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times m}$	mn	mn
	总　数	$(m+2)n(n+1)$	$(m+2)n^2$
	总 flop 数	$(m+2)n(2n+1)$	

表 9 多变量递推最小二乘算法的计算量

Table 9 The computational efficiency of the multivariable recursive least squares algorithm

变量	计算次序	乘法次数	加法次数
$\hat{\boldsymbol{\theta}}(t)$	$\boldsymbol{e}(t):=\boldsymbol{y}(t)-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^m$	mn	mn
	$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{L}(t)\boldsymbol{e}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times m}$	mn	mn
$\boldsymbol{L}(t)$	$\boldsymbol{\zeta}(t):=\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\in\mathbf{R}^n$	n^2	n^2-n
	$\boldsymbol{L}(t):=\boldsymbol{\zeta}(t)/[1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(t)]\in\mathbf{R}^n$	$2n$	n
$\boldsymbol{P}(t)$	$\boldsymbol{P}(t)=\boldsymbol{P}(t-1)-\boldsymbol{L}(t)\boldsymbol{\zeta}^{\mathrm{T}}(t)\in\mathbf{R}^{n\times n}$	n^2	n^2
	总　数	$2n^2+2mn+2n$	$2n^2+2mn$
	总 flop 数	$4n^2+4mn+2n$	

参考文献
References

[1] 丁锋. 系统辨识新论[M]. 北京:科学出版社,2012
DING Feng. System identification:New theory and methods[M]. Beijing:Science Press,2012

[2] 丁锋. 系统辨识(1):辨识导引[J]. 南京信息工程大学学报:自然科学版,2011,3(1):1-22
DING Feng. System identification. Part A:Introduction to the identification[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2011,3(1):1-22

[3] 丁锋. 系统辨识(2):系统描述的基本模型[J]. 南京信息工程大学学报:自然科学版,2011,3(2):97-117
DING Feng. System identification. Part B:Basic models for system description[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2011,3(2):97-117

[4] 丁锋. 系统辨识(3):辨识精度与辨识基本问题[J]. 南京信息工程大学学报:自然科学版,2011,3(3):193-226
DING Feng. System identification. Part C:Identification

accuracy and basic problems[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2011,3(3):193-226

[5] 丁锋.系统辨识(4):辅助模型辨识思想与方法[J].南京信息工程大学学报:自然科学版,2011,3(4):289-318
DING Feng. System identification. Part D:Auxiliary model identification idea and methods[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2011,3(4):289-318

[6] 丁锋.系统辨识(5):迭代搜索原理与辨识方法[J].南京信息工程大学学报:自然科学版,2011,3(6):481-510
DING Feng. System identification. Part E:Iterative search principle and identification methods[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2011,3(6):481-510

[7] 丁锋.系统辨识(6):多新息辨识理论与方法[J].南京信息工程大学学报:自然科学版,2012,4(1):1-28
DING Feng. System identification. Part F: Multi-innovation identification theory and methods[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2012,4(1):1-28

[8] 丁锋.系统辨识(7):递阶辨识原理与方法[J].南京信息工程大学学报:自然科学版,2012,4(2):97-124
DING Feng. System identification. Part G:Hierarchical identification principle and methods[J]. Journal of Nanjing University of Information Science & Technology:Natural Science Edition,2012,4(2):97-124

[9] 丁锋.系统辨识(8):耦合辨识概念与方法[J].南京信息工程大学学报:自然科学版,2012,4(3):193-212
DING F. System identification. Part H:Coupling identification concept and methods[J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition,2011,4(3):193-212

[10] Golub G H,Van Loan C F. Matrix computations[M],3rd Ed. Baltimore,MD:Johns Hopkins Univ Press,1996

[11] Liu Y J,Sheng J,Ding R F. Convergence of stochastic gradient algorithm for multivariable ARX-like systems [J]. Computers and Mathematics with Applications, 2010,59(8):2615-2627

[12] 丁锋,杨慧中,刘飞.弱条件下随机梯度算法性能分析[J].中国科学:E辑,2008,38(12):2173-2184

[13] Ding F, Yang H Z, Liu F. Performance analysis of stochastic gradient algorithms under weak conditions[J]. Science in China Series:Information Sciences,2008,51(9):1269-1280

[14] 丁锋,谢新民.传递函数阵子子模型参数递推估计:辅助模型方法[J].控制与决策,1991,6(6):447-452
DING Feng,XIE Xinmin. Recursive parameter estimation of transfer matrix sub-submodels: An auxiliary model method[J]. Control and Decision,1991,6(6):447-452

[15] 丁锋,谢新民.多变量系统的辅助模型辨识算法[J].清华大学学报:自然科学版,1992,32(4):100-106
DING Feng,XIE Xinmin. Auxiliary model identification method for multivariable systems[J]. Journal of Tsinghua University: Science and Technology Edition, 1992, 32(4):100-106

[16] 丁锋.多变量系统的辅助模型辨识方法的收敛性分析[J].控制理论与应用,1997,14(2):192-200
DING Feng. Convergence analysis of auxiliary model identification algorithms for multivariable systems[J]. Control Theory and Applications,1997,14(2):192-200

[17] Ding F,Chen T. Identification of dual-rate systems based on finite impulse response models[J]. International Journal of Adaptive Control and Signal Processing,2004,18(7):589-598

[18] Ding F,Chen T. Combined parameter and output estimation of dual-rate systems using an auxiliary model[J]. Automatica,2004,40(10):1739-1748

[19] Ding F,Chen T. Parameter estimation of dual-rate stochastic systems by using an output error method[J]. IEEE Transactions on Automatic Control,2005,50(9):1436-1441

[20] Ding F,Ding J. Least squares parameter estimation with irregularly missing data[J]. International Journal of Adaptive Control and Signal Processing,2010,24(7):540-553

[21] 丁锋,谢新民,方崇智.时变系统辨识的多新息方法[J].自动化学报,1996,22(1):85-91
DING Feng,XIE Xinmin,FANG Chongzhi. Multi-innovation identification methods for time-varying systems[J]. Acta Automatica Sinica,1996,22(1):85-91

[22] 丁锋,杨家本.衰减激励条件下确定性系统多新息算法的收敛性分析[J].清华大学学报:自然科学版,1998,38(9):111-115
DING Feng, YANG Jiaben. Convergence analysis of multi-innovation identification under attenuating excitation conditions for deterministic systems[J]. Journal of Tsinghua University:Science and Technology Edition,1998,38(9):111-115

[23] 丁锋,丁韬.时变多变量系统多新息投影算法的均方收敛性[J].湖北工学院学报,2001,16(4):16-20
DING Feng, DING Tao. Mean square convergence of multi-innovation projection algorithms for time-varying multivariable systems[J]. Journal of Hubei University of Technology,2001,16(4):16-20

[24] 丁锋,萧德云,丁韬.多新息随机梯度辨识方法[J].控制理论与应用,2003,20(6):870-874
DING Feng, XIAO Deyun, DING Tao. Multi-innovation stochastic gradient identification methods[J]. Control Theory and Application,2003,20(6):870-874

[25] Ding F,Chen T. Performance analysis of multi-innovation gradient type identification methods[J]. Automatica,2007,43(1):1-14

[26] Ding F,Chen H B,Li M. Multi-innovation least squares identification methods based on the auxiliary model for MISO systems[J]. Applied Mathematics and Computation,2007,187(2):658-668

[27] Ding F,Liu X P,Liu G. Auxiliary model based multi-in-

novation extended stochastic gradient parameter estimation with colored measurement noises[J]. Signal Processing,2009,89(10):1883-1890

[28] Liu Y J,Xiao Y S,Zhao X L. Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model[J]. Applied Mathematics and Computation,2009,215(4):1477-1483

[29] 于丽,刘艳君,丁锋. CARMA 模型多新息增广随机梯度参数估计算法的收敛性分析[J]. 系统工程与电子技术,2009,31(6):1446-1449
YU Li,LIU Yanjun,DING Feng. Convergence of multi-innovation extended stochastic gradient parameter estimation for CARMA models[J]. Systems Engineering and Electronics,2009,31(6):1446-1449

[30] Wang D Q,Ding F. Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems[J]. Digital Signal Processing,2010,20(3):750-762

[31] Ding F,Liu X P,Liu G. Multi-innovation least squares identification for system modeling[J]. IEEE Transactions on Systems,Man,and Cybernetics. Part B:Cybernetics,2010,40(3):767-778

[32] Ding F. Several multi-innovation identification methods [J]. Digital Signal Processing,2010,20(4):1027-1039

[33] Liu Y J,Yu L,Ding F. Multi-innovation extended stochastic gradient algorithm and its performance analysis [J]. Circuits,Systems and Signal Processing,2010,29(4):649-667

[34] Xie L,Yang H Z,Ding F. Modeling and identification for non-uniformly periodically sampled-data systems[J]. IET Control Theory and Applications,2010,4(5):784-794

[35] Ding F,Liu G,Liu X P. Parameter estimation with scarce measurements[J]. Automatica,2011,47(8):1646-1655

[36] 丁锋,杨家本. 大系统的递阶辨识[J]. 自动化学报,1999,25(5):647-654
DING Feng,YANG Jiaben. Hierarchical identification of large scale systems[J]. Acta Automatica Sinica,1999,25(5):647-654

[37] Ding F,Yang J B,Xu Y M. Convergence of hierarchical stochastic gradient identification for transfer function matrix models[J]. Control Theory and Applications,2001,18(6):949-953

[38] 丁锋,丁韬,杨家本,等. 衰减激励条件下递阶最小二乘辨识的均方收敛性[J]. 控制与决策,2002,17(1):6-10
DING Feng,DING Tao,YANG Jiaben,et al. Mean square convergence of hierarchical least squares identification under the attenuating excitation[J]. Control and Decision,2002,17(1):6-10

[39] 赵霞,姚郁,方强. 递阶辨识方法在转台伺服系统调试中的应用研究[J]. 控制理论与应用,2002,19(2):229-234
ZHAO Xia,YAO Yu,FANG Qiang. The study on the hierarchical identification method in the debugging of turntable servo systems[J]. Control Theory and Applications,2002,19(2):229-234

[40] 丁锋,陈通文,萧德云. 一般双率系统状态空间模型及其递阶辨识[J]. 自动化学报,2004,30(5):652-663
DING Feng,CHEN Tongwen,XIAO Deyun. State-space modeling and hierarchical identification for general dual-rate stochastic systems[J]. Acta Automatica Sinica,2004,30(5):652-663

[41] 丁锋,陈通文,萧德云. 非均匀周期采样多率系统的一种递阶辨识方法[J]. 电子学报,2004,32(9):1414-1420
DING Feng,CHEN Tongwen,XIAO Deyun. Hierarchical identification of non-uniformly periodically sampled multirate systems[J]. Acta Electronica Sinica,2004,32(9):1414-1420

[42] 丁锋,萧德云. 多变量系统状态空间模型的递阶辨识[J]. 控制与决策,2005,20(8):848-853,859
DING Feng,XIAO Deyun. Hierarchical identification of state space models for multivariable systems[J]. Control and Decision,2005,20(8):848-853,859

[43] Ding F,Chen T. Hierarchical gradient-based identification of multivariable discrete-time systems[J]. Automatica,2005,41(2):315-325

[44] Ding F,Chen T. Hierarchical least squares identification methods for multivariable systems[J]. IEEE Transactions on Automatic Control,2005,50(3):397-402

[45] Ding F,Chen T. Hierarchical identification of lifted state-space models for general dual-rate systems[J]. IEEE Transactions on Circuits and Systems. I:Regular Papers,2005,52(6):1179-1187

[46] Wang L Y,Ding F,Liu X P. Consistency of HLS estimation algorithms for MIMO ARX-like systems[J]. Applied Mathematics and Computation,2007,190(2):1081-1093

[47] Han H Q,Xie L,Ding F,et al. Hierarchical least squares based iterative identification for multivariable systems with moving average noises[J]. Mathematical and Computer Modelling,2010,51(9/10):1213-1220

[48] Xiang L L,Xie L B,Ding R F. Hierarchical least squares algorithms for single-input multiple-output systems based on the auxiliary model[J]. Mathematical and Computer Modelling,2010,52(5/6):918-924

[49] Zhang Z N,Ding F,Liu X G. Hierarchical gradient based iterative parameter estimation algorithm for multivariable output error moving average systems[J]. Computers and Mathematics with Applications,2011,61(3):672-682

[50] Ding J,Ding F,Liu X P,et al. Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data[J]. IEEE Transactions on Automatic Control,2011,56(11):2677-2683

[51] Ding F,Liu G,Liu X P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems[J]. IEEE Transactions on Automatic Control,2010,55(8):1976-1981

Computational efficiency of the identification methods. Part A:Recursive algorithms

DING Feng[1,2,3]
1 School of Internet of Things Engineering,Jiangnan University,Wuxi 214122
2 Control Science and Engineering Research Center,Jiangnan University,Wuxi 214122
3 Key Laboratory of Advanced Process Control for Light Industry(Ministry of Education),Jiangnan University,Wuxi 214122

Abstract The amount of the calculation of an algorithm may be expressed by the number of multiplication and addition operations (one division is treated as a multiplication,one subtraction treated as an addition). A multiplication or an addition operation is called a flop,i. e. ,a floating-point operation. This is the first of three serial papers \`Computational efficiency of the identification methods',which focuses on the computational efficiency of the recursive algorithms,including the flops of the vector and matrix operations,and the minimum flops of the stochastic gradient identification algorithm,the least squares identification algorithm,the recursive least squares identification algorithm for linear regression systems,multivariate linear regression systems and multivariable systems.

Key words recursive identification;iterative identification;parameter estimation;FIR model;equation error model;CAR model;CARMA model;CARAR model;CARARMA model;output error model;OEMA model;OEAR model;auxiliary model identification;multi-innovation identification;hierarchical identification;coupled identification