

## Volterra 级数模型的非线性压缩测量辨识算法\*

邱 棚<sup>1,2</sup>, 姚旭日<sup>1</sup>, 李鸣谦<sup>1,2</sup>, 翟光杰<sup>1</sup>

(1. 中国科学院国家空间科学中心, 北京 100190; 2. 中国科学院大学, 北京 100049)

**摘要:**针对非线性系统的辨识问题,提出了非线性压缩测量辨识算法,且推导出了一符合压缩感知测量准则的测量模型。相比递归最小二乘法,该方法极大地减少了所需的测量数,使得高阶 Volterra 级数辨识成为可能。此外,还分析了实际应用中的各项因素对辨识准确性的影响,如信号稀疏度、测量噪声、测量矩阵形式等。

**关键词:**系统辨识;非线性系统;压缩感知;Volterra 级数;正交匹配追踪

**中图分类号:**TN95 **文献标志码:**A **文章编号:**1001-2486(2020)01-125-08

## Nonlinear compressed measurement identification based on Volterra series

QIU Peng<sup>1,2</sup>, YAO Xuri<sup>1</sup>, LI Mingqian<sup>1,2</sup>, ZHAI Guangjie<sup>1</sup>

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** For the identification problem of nonlinear systems, the accuracy and stability of the nonlinear compression measurement identification algorithm were proved in the simulation experiment, and the complete signal was obtained accurately only by using constant multiple measurement times of the signal sparsity. Compared with the least square method, the proposed algorithm has greatly reduced the needed measurements, therefore, it is possible for the identification of high-order Volterra series. Furthermore, the influence of all factors on the accuracy of system identification was analyzed, such as signal sparsity, measurement noise, measurement matrix form, etc.

**Keywords:** system identification; nonlinear system; compressed sensing; Volterra series; orthogonal matching pursuit

一直以来,关于非线性系统的研究吸引了来自各个领域学者的关注。特别是在工程领域中,理想的线性系统是几乎不存在的,大部分需要面对的都是非线性系统。一般认为,输出与输入不成比例的系统均是非线性的,如失真、自激振荡、分叉、混沌等。本文重点研究弱非线性系统,比如功率放大器<sup>[1]</sup>、扬声器<sup>[2]</sup>、生物过程<sup>[3]</sup>均属此类。对于此类非线性系统,其模型是很难通过理论推导得到的。因而,需要通过输入输出信号的辨识得到系统模型。

在系统辨识中,首要考虑的问题是如何表示非线性系统。到目前为止,还没有一个通用的非线性系统表示方法。非线性系统的模型可以从以下几个方式获得:获得系统的频域等价模型;将非线性环节表示为多个线性环节的卷积;将非线性系统展开为线性形式,再利用线性的辨识方法

获得非线性系统的近似等。本文采用后一种方法,利用多项式展开的方法来表示非线性对象。根据泰勒公式可知,如果系统的非线性环节足够平滑并且没有历史记忆,那么就可将系统的输出表示为输入信号的多项式展开。但是,如果系统是有记忆的,就需要用 Volterra 级数来描述<sup>[4]</sup>。本文选用  $p$  阶、记忆长度为  $L$  的 Volterra 核作为基础,并利用多项式回归的方法得到核的参数,从而使展开项的和尽可能地接近原函数。近年来,Volterra 级数以及其他多项式展开的方式在人脸识别<sup>[5]</sup>、语音识别<sup>[6]</sup>、文字识别<sup>[7]</sup>等非线性问题上都取得了成功。

系统辨识的第二个问题是如何准确估计出模型中的参数。鉴于 Volterra 级数中的未知参数可以表示为输入输出的线性组合,经典的最小二乘(Least-Square, LS)方法就可辨识得到<sup>[8]</sup>。但是,

\* 收稿日期:2018-10-01

基金项目:国家自然科学基金资助项目(61605218);中国科学院国防创新基金资助项目(CXJJ-17S023)

作者简介:邱棚(1991—),男,北京人,博士研究生,E-mail:qpeng0504@163.com;

翟光杰(通信作者),男,教授,博士,博士生导师,E-mail:gjzhai@nssc.ac.cn

此方法会碰到“维数灾难”的问题,即待辨识的参数数量将随着模型阶数的增加而呈指数增长。对于 LS 方法而言,就需要远大于  $\Theta(L^p)$  次的数据,辨识结果才能收敛。这不仅增加了辨识的成本,还会增大计算复杂度。一个可能的方法是,将该问题视为核回归问题<sup>[9-10]</sup>。然而,Volterra 级数的高阶参数是稀疏度的,也就是说待辨识参数中只有个别位置是非零的。不过,多项式的阶数、记忆长度以及非零项的位置和个数都是未知的,所以无法在辨识之前选定合适的数值。此外,在实际系统中,核回归的方法由于输入输出数据存在误差将导致过拟合,使得辨识结果变差。考虑到参数的稀疏性,有学者提出采用稀疏估计器来解决过拟合的问题,典型的如非负参数推断(Non-Negative Garrote, NNG)<sup>[11]</sup>和递归最小二乘(Recursive Least Squares, RLS)估计器。文献[18]给出了针对稀疏 Volterra 级数的 EM-RLS 算法,对参数的期望和分布进行计算,从而得到准确的辨识结果。另一个常见的稀疏估计器叫作最小绝对收缩和选择算子(Least-Absolute Shrinkage and Selection Operator, LASSO)<sup>[13]</sup>。文献[14]基于该算法,提出了针对 Volterra 级数的 wLASSO (weight LASSO)估计器,并取得了良好的效果,而且 wLASSO 的性能已经在文献[15]之中分析过了。除了利用稀疏性解决“维数灾难”问题的方法以外,还有利用核的对称性以减少待辨识参数的个数<sup>[8]</sup>。另外,还可以利用正交三角分解的方法来减少辨识计算量<sup>[16]</sup>。

近年来,Donoho<sup>[17]</sup>、Candès 等<sup>[18]</sup>提出的压缩感知理论是处理信号稀疏性的一种高效方法。根据采样定理可知,对任何信号的测量都要求采样频率是信号带宽的 2 倍以上。然而,压缩感知理论指出,对稀疏信号的测量需要关注的是稀疏度,而不是信号带宽。该定理目前已经被大量应用于有关成像的各个领域中,如高速流显微镜成像<sup>[19]</sup>、核磁共振成像<sup>[20]</sup>、合成孔径雷达<sup>[21]</sup>、毫米波雷达信号处理<sup>[22]</sup>等。Volterra 级数中的一阶核可以视为一维的信号,其二阶核也可视为二维的信号。依此类推,级数的各阶核参数均可以视作高维信号,而且这些信号均是极度稀疏的。因而,利用压缩感知理论,可以极大地提高测量的效率。

## 1 多项式级数展开

本节主要介绍非线性系统的 Volterra 级数展开方式。假设一个离散非线性时不变系统的输出可表示为:

$$y(n) = f(x(n), x(n-1), \dots, x(n-l)) \quad (1)$$

其中,  $x(n)$  代表系统在时刻  $n$  的输入,而  $f(x)$  则是一个非线性函数。上式表示,系统的输出与当前时刻之前的所有时刻都相关,即系统是有无限记忆的。但是,对于大部分系统而言,系统的响应都会随着时间的推移而衰减。所以,在此假设系统的记忆是有限且合理的,也就是系统的输出只与前  $L$  个时刻的输入有关。本文只考虑由死区、饱和、谐波等导致的信号失真等弱非线性系统。理论上,Volterra 级数是一个无穷级数,而本文允许展开级数与原系统存在截断误差,所以系统的输出就可以利用  $P$  阶的 Volterra 级数来表示:

$$y(n) = \sum_{p=0}^P H_p[X_1(n)] + v(n) \quad (2)$$

式中:  $v(n)$  包含截断误差以及测量噪声,而且该噪声是一个零均值并且独立于输入信号的白噪声;  $H_p[X_1(n)]$  表示第  $p$  阶的 Volterra 核,其具体形式如下:

$$H_p[X_1(n)] = \sum_{k_1=0}^{L-1} \dots \sum_{k_p=0}^{L-1} h_p(k_1, \dots, k_p) \prod_{i=1}^p x(n-k_i) \quad (3)$$

式中,  $L$  代表 Volterra 级数的记忆长度,而且各阶核的记忆长度是一致的。针对 Volterra 级数展开的收敛域以及使用范围已经被广泛研究,可以参考文献[8]。本文的目的是利用系统的输入输出数据,以辨识得到可以准确描述系统的各阶核参数  $h_p(k_1, \dots, k_p)$ 。

## 2 非线性压缩测量辨识

非线性压缩测量辨识算法将压缩感知理论引入对非线性系统的辨识之中,并利用测量的方法对模型参数进行观测。在测量的框架下,首先要确定被测信号的形式。在系统辨识中,被测信号包含了系统的全部特征,每一个信号都可以唯一地表示一个系统。对于线性系统来说,可以选择系统的脉冲响应序列来表示,该信号包含了线性系统的全部动态和静态特征。那么,对于非线性系统来说,是否存在一个信号是可以唯一地表达系统的呢? 本文采用 Volterra 级数的各阶核来表示。另外,在测量的角度下,可以认为表示系统的信号是客观存在的,只不过需要特殊的观测方法才能够得到,或者是需要通过间接的方法来对其测量。本文的重点在于介绍测量在辨识非线性系统中的思想,因而只考虑二阶 Volterra 级数的辨识问题,而更高阶系统可以依此推广。由于忽略了高阶核,所以测量结果与原信号存在一定偏差,

而这个偏差从测量的角度来看,是由于测量工具的局限性导致的。这就如同用米尺去测量毫米级的物体一般,一定会存在误差。

假设非线性系统展开为二阶 Volterra 级数,该系统输出可以表示为:

$$y_k = h + \sum_{t=0}^{L-1} h(t)u(k-t) + \sum_{t_1=0}^{L-1} \sum_{t_2=0}^{L-1} h(t_1, t_2)u(k-t_1)u(k-t_2) \quad (4)$$

根据上文对 Volterra 级数的介绍可知,一个非线性系统的输出可以写为多维卷积的形式。对于一阶 Volterra 核,可以将一维卷积改写为向量相乘的形式,即:

$$\mathbf{y}_k^1 = (u_k, u_{k-1}, \dots, u_{k-L+1}) \cdot (h_0^1, h_1^1, \dots, h_{L-1}^1)^T \quad (5)$$

式中:  $\mathbf{y}_k^1$  表示一阶核第  $k$  时刻的系统输出;  $u_k$  表示第  $k$  时刻的系统输入;  $h_i^1$  表示一阶核第  $i$  时刻的脉冲响应。如果考虑  $M$  个时刻的系统输出,此时可以将上式写为矩阵形式  $\mathbf{Y}^{(1)} = \mathbf{U}^{(1)} \mathbf{H}_1$ , 则:

$$\begin{bmatrix} \mathbf{y}_k^1 \\ \mathbf{y}_{k+1}^1 \\ \vdots \\ \mathbf{y}_{k+M-1}^1 \end{bmatrix} = \begin{bmatrix} u_k & u_{k-1} & \cdots & u_{k-L+1} \\ u_{k+1} & u_k & \cdots & u_{k-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k-1+M} & u_{k-2+M} & \cdots & u_{k-L+M} \end{bmatrix} \begin{bmatrix} h_0^1 \\ h_1^1 \\ \vdots \\ h_{L-1}^1 \end{bmatrix} \quad (6)$$

根据式(2)可知,二阶 Volterra 核的输出可以写为线性二次型的形式,即:

$$\mathbf{y}_k^2 = \begin{bmatrix} u_k \\ u_{k-1} \\ \vdots \\ u_{k-L+1} \end{bmatrix}^T \begin{bmatrix} h_{0,0}^2 & h_{0,1}^2 & \cdots & h_{0,L-1}^2 \\ h_{1,0}^2 & h_{1,1}^2 & \cdots & h_{1,L-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ h_{L-1,0}^2 & h_{L-1,1}^2 & \cdots & h_{L-1,L-1}^2 \end{bmatrix} \begin{bmatrix} u_k \\ u_{k-1} \\ \vdots \\ u_{k-L+1} \end{bmatrix} \quad (7)$$

式中:  $\mathbf{y}_k^2$  表示二阶核在第  $k$  时刻的输出;  $h_{i,j}^2$  表示二阶核对由第  $i, j$  时刻的输入所引起的系统响应大小。此时,将  $h_{i,j}^2$  逐行展开成一个列向量,式(6)进一步改写为两个向量相乘的形式。同样,考虑  $M$  个时刻的系统输出,二阶核的输出也可以写为矩阵形式  $\mathbf{Y}^{(2)} = \mathbf{U}^{(2)} \mathbf{H}_2$ , 则:

$$\begin{bmatrix} \mathbf{y}_k^2 \\ \mathbf{y}_{k+1}^2 \\ \vdots \\ \mathbf{y}_{k+M-1}^2 \end{bmatrix} = \begin{bmatrix} u_k^2 & \cdots & u_{k-1}u_k & \cdots & u_{k-L+1}^2 \\ u_{k+1}^2 & \cdots & u_k u_{k+1} & \cdots & u_{k-L+2}^2 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{k+M-1}^2 & \cdots & u_{k+M-1}u_{k+M} & \cdots & u_{k-L+M}^2 \end{bmatrix} \begin{bmatrix} h_{0,0}^2 \\ h_{0,1}^2 \\ \vdots \\ h_{1,0}^2 \\ \vdots \\ h_{L-1,L-1}^2 \end{bmatrix} \quad (8)$$

此时,将一阶核和二阶核先后排列成一个一维列向量,那么非线性系统输出可以表示为:

$$\mathbf{Y} = \mathbf{Y}^{(0)} + \mathbf{Y}^{(1)} + \mathbf{Y}^{(2)} = [\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}][\mathbf{H}_c, \mathbf{H}_1, \mathbf{H}_2]^T \quad (9)$$

式中,  $\mathbf{U}^{(0)}$  是一个  $M \times 1$  的常数向量,  $\mathbf{H}_c$  是一个  $1 \times 1$  的常数向量。  $\mathbf{H} = [\mathbf{H}_c, \mathbf{H}_1, \mathbf{H}_2]^T$  是待测量信号,该信号长度  $N = L^2 + L$ 。不过,由于二阶核是关于对角线对称的,因此可以将相同项合并,即信号长度缩减为  $N = (L^2 + 3L + 2)/2$ ;  $\mathbf{U} = [\mathbf{U}^{(0)}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}]$  是测量矩阵,该矩阵中的每一行都是一个测量向量,用于对信号  $\mathbf{H}$  进行一次观测。因而,矩阵  $\mathbf{U}$  的性质直接决定了测量的准确性。压缩感知理论给出了准确测量的要求,即只有当测量矩阵满足有限等距约束 (Restricted Isometry Property, RIP) 性质时,才能保证测量是无损的。学者已经证明一般常见的独立同分布的随机矩阵都满足 RIP 性质。不过,对于本文中使用的复杂随机矩阵,要想得到 RIP 性质的充要条件是十分困难的。因而,可以通过证明得到满足 RIP 性质的充分条件来得到对测量矩阵的要求。文献[23]中给出了当输入信号是符合均匀分布的随机信号时的定理如下:

**引理**<sup>[23]</sup> 假设给定输入序列的各项  $\{u_i\}_{i=L-1}^N$  是随机的、从均匀分布  $\mathbf{U} \in [-1, 1]$  中取出的,定义信号长度  $N = (L+1)(L+2)/2$ 。当 Volterra 核的记忆长度  $L \geq 7$ , 并且观测次数  $M \geq 160$ , 那么就有如下推论。对于任意  $S$  阶有限等距常数  $\delta_S \in (0, 1)$  和参数  $\gamma \in (0, 1)$ , 只要测量数  $M \geq S^2 \lg L 5C / (1 - \gamma) \delta_S^2$ , 那么测量矩阵  $\mathbf{U}$  符合  $S$  阶 RIP 性质的可能性超过  $1 - \exp\left(-\frac{\gamma \delta_S^2 N}{C S^2}\right)$ , 其中常数  $C = 2835$ 。

该引理表明,当利用符合均匀分布的测量矩阵对信号进行测量时,如果测量数是与  $S^2 \lg N$  在同阶大小时,恢复算法是有极大的概率从观测值中重建出被测信号的。不过,此处只是给出了均匀分布的测量准则,而且该准则只是给出了一个相对宽泛的下限。此外,如果系统输入选取的是其他随机分布,就需要重新推导 RIP 定理。因而,本文尝试间接地给出测量矩阵需要满足的性质。根据 RIP 定理的定义可知,RIP 性质描述的是从测量矩阵中任意抽取任意列组成的子矩阵能否完整地保存原信号的能量,即不对原信号过分放大也不过衰减。因而,测量矩阵的子矩阵满足 RIP 性质是一个必要条件。利用圆盘定理来解释,可以认为 RIP 性质要求测量矩阵的 Gramian

矩阵  $\mathbf{R} = \mathbf{U}^T \mathbf{U}$  的特征值在  $[1 - \delta_s, 1 + \delta_s]$  范围内。概括来说,要想保证测量是准确,就需要满足以下三个条件:

- ①  $E[u_i] = 0, \mathbf{U} = [u_1^T, u_2^T, \dots, u_M^T]^T$ ;
- ②  $E[R_{i,i}] = 1, i = 1, \dots, N$ ;
- ③  $E[R_{i,j}] = 0, i, j = 1, \dots, N, i \neq j$ 。

上述条件要求测量向量只与自身线性相关,而且要与其他测量向量线性无关。因为系统输入是随机向量,因而很容易证明一阶核的测量向量满足该定理。但是,由于二阶核的测量向量是输入向量的非线性组合,因而二阶核测量向量的性质不容易得到证明。但是,可以通过重新排列将原测量矩阵改写为多个 Toeplitz 矩阵相连的形式,如式(10)所示。式中  $\tilde{\mathbf{U}}_0^{(2)}$  的下标代表  $i, j$  的差,即表示两次输入之间的时间间隔  $|i - j|$ 。那么,对于系统记忆长度为  $L$  的二阶核,可以将测量矩阵改写为  $L - 1$  个 Toeplitz 矩阵块。此时,可以单独考虑每个矩阵块的性质。比如,此时很容易计算得到各循环矩阵的方差期望,因为每一个循环矩阵都是结构化的,而且其特性都取决于第一行向量。特别是, RIP 性质要求其任意子矩阵均满足,因而取每一个循环矩阵 RIP 条件的并集,就可以得到整个测量矩阵 RIP 性质的一个必要条件。此外,对于更高阶的 Volterra 核,同样可以按照本文所述方法排列,即将核按照其对角线排列。

$$\mathbf{U} = [\tilde{\mathbf{U}}^{(1)} \quad \tilde{\mathbf{U}}_0^{(2)} \quad \tilde{\mathbf{U}}_1^{(2)} \quad \dots \quad \tilde{\mathbf{U}}_{L-1}^{(2)} \quad \tilde{\mathbf{U}}^{(c)}]$$

其中

$$\left\{ \begin{aligned} \tilde{\mathbf{U}}_0^{(2)} &= \begin{bmatrix} u_k^2 & u_{k-1}^2 & \dots & u_{k-L+1}^2 \\ u_{k+1}^2 & u_k^2 & \dots & u_{k-L+2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ u_{k+M-1}^2 & u_{k+M}^2 & \dots & u_{k-L+M}^2 \end{bmatrix} \\ \tilde{\mathbf{U}}_1^{(2)} &= \begin{bmatrix} u_k u_{k-1} & \dots & u_{k-L+1} u_{k-L} \\ u_{k+1} u_k & \dots & u_{k-L+2} u_{k-L+1} \\ \vdots & \ddots & \vdots \\ u_{k+M-1} u_{k+M-2} & \dots & u_{k-L+M} u_{k-L+M-1} \end{bmatrix} \\ \tilde{\mathbf{U}}_2^{(2)} &= \begin{bmatrix} u_k u_{k-2} & \dots & u_{k-L+2} u_{k-L} \\ u_{k+1} u_{k-1} & \dots & u_{k-L+3} u_{k-L+1} \\ \vdots & \ddots & \vdots \\ u_{k+M-1} u_{k+M-3} & \dots & u_{k-L+M+2} u_{k-L+M-1} \end{bmatrix} \\ \tilde{\mathbf{U}}_{L-1}^{(2)} &= \begin{bmatrix} u_k u_{k-L+1} \\ u_{k+1} u_{k-L+2} \\ \vdots \\ u_{k+M} u_{k-L+M} \end{bmatrix} \end{aligned} \right. \quad (10)$$

假设输入信号服从标准分布  $N(0,1)$ ,那么可以计算得出除  $\tilde{\mathbf{U}}_0^{(2)}$  中各列的期望为 1,其他循环矩阵各行的期望为 0。此时只需将  $\tilde{\mathbf{U}}_0^{(2)}$  矩阵的每项都减去 1,测量矩阵就可满足条件 ①,即  $\hat{\mathbf{U}}_0^{(2)} = \tilde{\mathbf{U}}_0^{(2)} - \mathbf{1}$ 。这相当于给每次的观测值多减去了  $C_h = \sum_{i=0}^{L-1} h_{i,i}$ 。对于一个确定的系统,  $C_h$  的大小是一定的,可以将其影响并入  $H_0$  中。为了满足条件 ② 和条件 ③,需要对测量矩阵进行归一化。因为  $E[u^2] = 1, E[u^4] = 3$ ,所以对矩阵的归一化结果如下:

$$\tilde{\mathbf{U}} = \begin{bmatrix} \sqrt{\frac{1}{M}} \tilde{\mathbf{U}}^{(1)} & \sqrt{\frac{1}{2M}} (\tilde{\mathbf{U}}_0^{(2)} - \mathbf{1}) & \sqrt{\frac{1}{M}} \tilde{\mathbf{U}}_1^{(2)} \\ \dots & \dots & \dots \\ \sqrt{\frac{1}{M}} \tilde{\mathbf{U}}_{L-1}^{(2)} & \sqrt{\frac{1}{M}} \tilde{\mathbf{U}}^{(c)} \end{bmatrix}$$

此时,测量矩阵  $\tilde{\mathbf{U}}$  同时满足上述三个条件。由于更改了测量向量的大小,因而重建后的信号与原信号不一致,需要通过如下方式映射,得到原信号:

$$\mathbf{H} = \begin{bmatrix} \sqrt{\frac{1}{M}} \tilde{\mathbf{H}}^{(1)} & \sqrt{\frac{1}{2M}} \tilde{\mathbf{H}}_0^{(2)} & \sqrt{\frac{1}{M}} \tilde{\mathbf{H}}_1^{(2)} & \dots \\ \sqrt{\frac{1}{M}} \tilde{\mathbf{H}}_{L-1}^{(2)} & \sqrt{\frac{1}{M}} \tilde{\mathbf{H}}^{(c)} - \sqrt{\frac{1}{2M}} \sum \tilde{\mathbf{H}}_0^{(2)} \end{bmatrix}^T$$

### 3 仿真

本节将测试非线性测量辨识算法在实际非线性系统中的辨识结果与真实系统的接近程度。从测量的角度来看,对测量结果影响最大的三个因素就是测量值的个数、测量噪声以及测量向量的形式。因而,本文将测试这些参数对测量结果的影响。

本文选用一个常见的非线性系统,即 Wiener-Hammerstein 模型作为辨识对象。该模型是由一个线性环节、一个非线性环节和另一个线性环节串联组成,如图 1 所示。

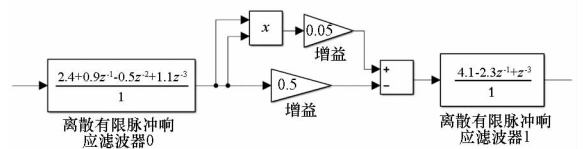


图 1 Simulink 模型

Fig. 1 Simulink model

本文利用 Simulink 搭建了一个离散系统的仿真环境,其中线性环节为离散脉冲序列(如图 1 所示),静态的非线性环节为  $Y(x) = -0.5x +$

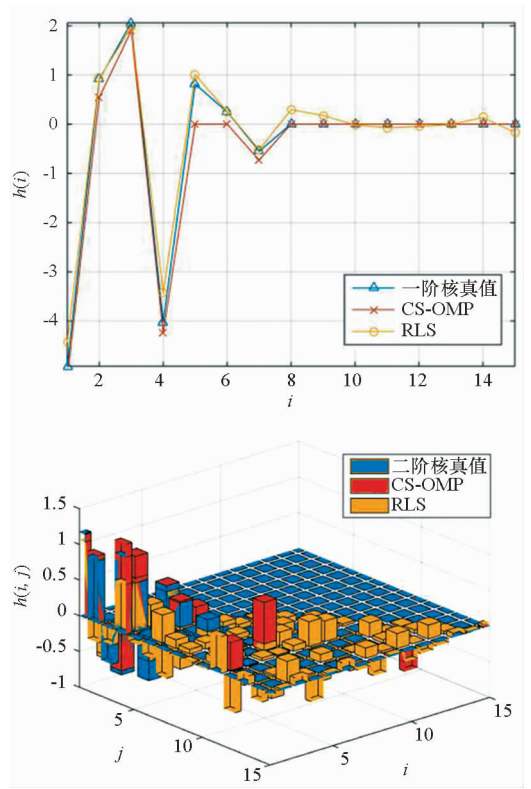
$0.05x^2$ 。为了接近真实环境,还在系统的输出上叠加了服从高斯分布的噪声  $\varepsilon(t) \sim N(0, 0.1)$ 。鉴于系统的记忆长度是未知的,设定一个足够长的记忆长度  $L = 15$ ,那么辨识信号总长度  $N = 136$ 。为了满足非线性压缩测量算法的要求,选定系统的输入信号为一系列服从高斯分布的随机信号  $u(n) \sim N(0, 1)$ 。本文采用了 100 次蒙特卡洛测试来验证算法的准确性。考虑到测量数  $M$  是小于等于信号长度  $N$  的,本文选用基于最小二范数的递推最小二乘辨识算法作为对比。对于辨识结果,本文从平均拟合程度 (Average Fit Rate, AFR) 和最小均方误差 (Mean Square Error, MSE) 两个角度来评价原信号  $\mathbf{H}$  和辨识得到的信号  $\mathbf{H}^*$  的接近程度,  $\mathbf{H}_i^*$  表示第  $i$  次实验的估计结果,  $\bar{\mathbf{H}}_i^*$  表示 100 次实验结果的平均:

$$AFR = E \left[ 1 - \frac{\|\mathbf{H} - \mathbf{H}_i^*\|_2}{\|\mathbf{H} - \bar{\mathbf{H}}_i^*\|_2} \right] \times 100\% \quad (11)$$

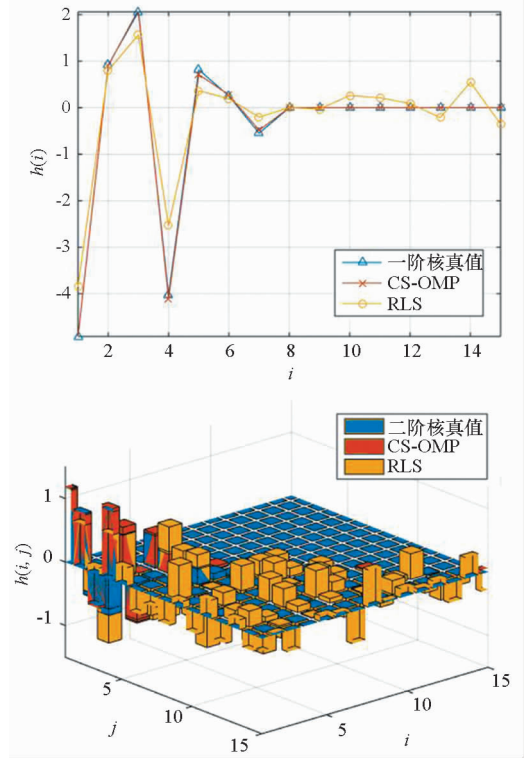
$$MSE = E [\|\mathbf{H} - \mathbf{H}_i^*\|_2^2] \quad (12)$$

图 2 呈现了两种辨识算法在两种测量数下的辨识结果。其中,一阶核在图 2(a) 中采用了折线来表示,而二阶核则采用了三维柱状图来表示。本文定义辨识过程中采集到的系统输出个数为测量数  $M$ ,又定义测量数  $M$  与信号长度  $N$  的比率为测量比率 (Measurement Ratio, MR)。图 2(a) 是在 MR 为 40% 的情况下的辨识结果,图中上半部分展现的是 Volterra 级数一阶核的测量结果和原信号,图 2(a) 下半部分展现的是二阶核的测量结果以及原信号。其中,蓝色代表原信号,黄色代表 RLS 的辨识结果,红色代表非线性 CMI 的辨识结果。图 2(b) 是在 MR 为 60% 的情况下的辨识结果,图 2(b) 中定义同上。从这两个图中可以看到,本文提出的算法相比 RLS 算法辨识到的一阶核和二阶核的准确性有很大程度的提高。由于原信号是稀疏的,也就是只有靠近原点附近的项是非零的,非线性 CMI 方法的测量结果基本符合了这一特点,而 RLS 算法却将原信号能量散布在了各项上。而且,随着测量率的提高,可以明显地看到非线性 CMI 方法的测量准确性有了进一步的提高。

本文仿真了测量率从 10% 到 200% 的所有情况,并将测试结果绘制在图 3 中。从图中可以看到,当测量率达到 66% 时,非线性 CMI 的拟合程度已经达到 94.13%,而递推最小二乘算法的测量率达到 177% 时,其平均拟合程度才达到同样水平。另外,当测量数大小接近于信号长度时,RLS 法出现了较大的偏差。这主要是由于观测存



(a) MR = 40%



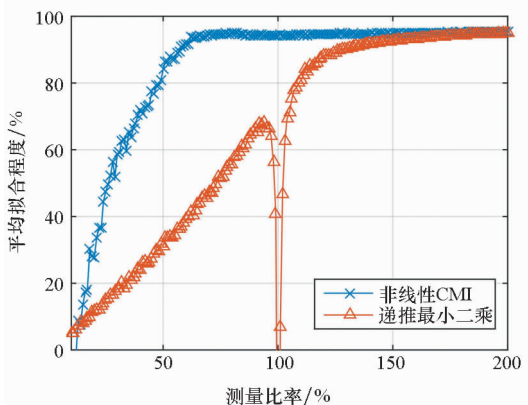
(b) MR = 60%

图 2 NCMi 和 RLS 辨识得到的 Volterra 核  
Fig. 2 Volterra core identified by NCMi and RLS

在误差,而当测量矩阵接近方阵时,RLS 算法的解空间越来越小,直到被压缩到唯一解。此时,所有观测噪声都被纳入系统模型之中,所以其偏差也就很大了。虽然,RLS 算法随着测量数的继续增

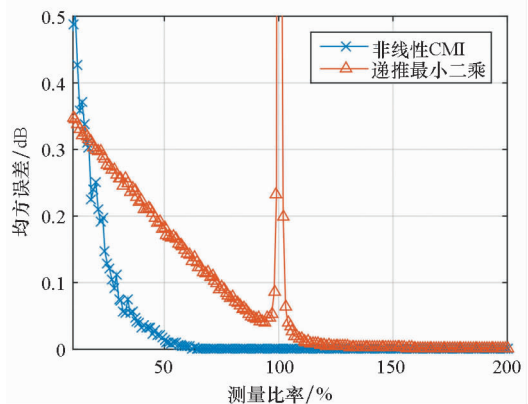


多,其结果的准确性也继续提高,不过最终也只能达到与非线性 CMI 方法同样的精度。也就是说,在该条件下,非线性 CMI 方法只需要 90 次测量就可以得到原信号,相比 RLS 能够节约大概 150 次测量值。



(a) 平均拟合程度与测量比率的关系

(a) AFR of identification results under different MR



(b) 均方误差与测量比率的关系

(b) MSE of Identification results under different MR

图 3 测量比率对 NCMI 和 RLS 的辨识影响

Fig. 3 Influence of NCMI and RLS identification results by MR

非线性 CMI 的测量结果在测量数达到 90 个时,就达到了最佳的测量结果。那么,如何确定这个数量,从而避免测量的浪费呢?在此,定义使得测量结果达到最佳的最小所需测量数为最小测量数。根据定理可知,最小测量数应该与稀疏度和信号长度有关,而稀疏度的影响远超过信号长度。因此,本文假设 Volterra 核的记忆长度为 30,阶数为 2,即信号总长度  $N$  为 496。其中,非零项随机地分布在信号上。图 4 展示了针对不同稀疏度信号的最小测量数。根据定理可知,最小测量数应该与稀疏度的二次方成比例。但是,在线性系统中,该数值与稀疏度呈线性关系。因而,本文利用直线和抛物线分别对结果进行拟合,并且计算了两种方法拟合结果与实际值的接近程度。其

中,抛物线的平均接近程度为 89.3%,而直线拟合的平均接近程度为 77.9%。由此可见,最小测量数与稀疏度的二次方是同阶的。换句话说,非线性 CMI 方法在测量过程中对信号的非零项个数更敏感。

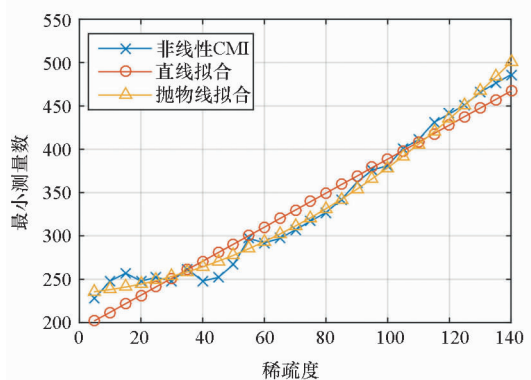


图 4 稀疏度对最小测量数的影响

Fig. 4 Influence of sparsity on minimum measurement number

在测量中,噪声是不可避免的,因而在包含测量噪声的条件下能否稳定地测量是至关重要的。对于测量算法来说,测量噪声的绝对大小是没有意义的,相对大小才是最重要的。因此,本文采用信噪比 (Signal Noise Rate, SNR) 作为指标,以测试噪声对算法的影响。图 5 展现了两种算法在不同噪声大小下的辨识准确度。可以看到,非线性 CMI 方法对噪声具有一定的鲁棒性,及时噪声达到 5 dB 时,信号的平均拟合程度还能保持在 69.3% 的水平。

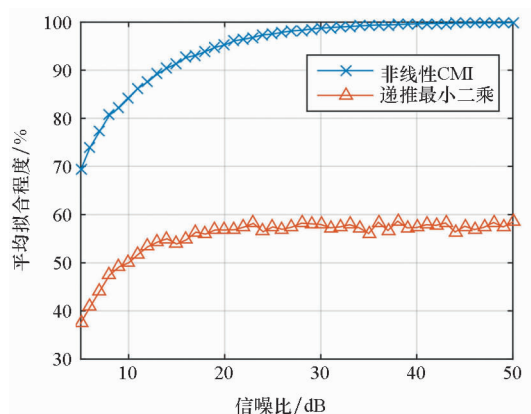


图 5 测量噪声对辨识结果的影响

Fig. 5 Influence of measurement noise on identification results

除了测量数和测量噪声以外,还有一个重要因素直接影响测量结果,那就是测量向量的形式。这是因为测量向量的形式直接决定了每次测量的方式以及测量获得的信息多少,或者说压缩感知

正是凭借着精心设计的测量向量,从而实现高效采样。由前文可知,选用的输入信号是随机地从一个高斯分布中取出的。由于一阶核的测量向量是系统的输入序列,因而一阶核的测量向量是服从高斯分布的。对于二阶核来说,可将核参数分为两类,其中一类是两个不同时刻的输入序列相乘。由于两个服从高斯分布的随机变量相乘后仍然服从高斯分布,所以第一类的测量向量仍然服从高斯分布。第二类是两个相同时刻的输入序列相乘,因而观测对角线信号的测量向量就不再服从高斯分布,而是服从卡方分布。图 6 比较了这两种分布对同一信号的测量结果。可以看出,虽然两个分布的最佳测量结果是一致的,但是高斯分布比卡方分布的最小测量数要少得多。这主要是因为卡方分布的各项均是非负的,也就是服从该分布的测量空间减少了一半,那么肯定就需要更多的观测才能达到同样的效果。

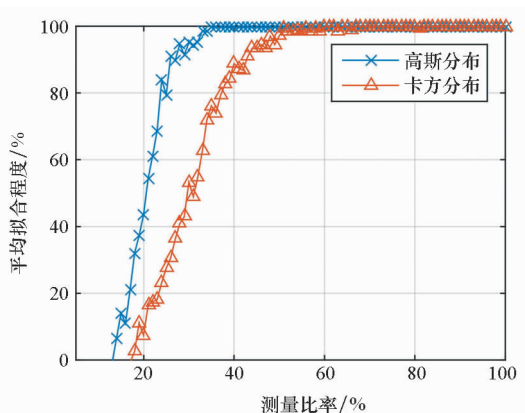


图 6 测量向量的分布对测量准确性的影响

Fig. 6 Influence of measurement vector distribution on accuracy

既然,Volterra 二阶核对角线元素的测量服从卡方分布,那么是否可以通过修改测量向量,从而提高测量效率呢?为了解决非负的问题,一个直接的方法是,将测量向量的各项减去常数,使得其均值为零。另外,如果某一行在测量中所占比例过大,那么也会影响到其他项的测量。因而,再对整个测量矩阵的每列分别做归一化。相应地,在获得测量结果后,要将这些修改复原才能得到原信号。图 7 展示了没有归一化测量向量之前的测量结果,以及归一化测量向量之后的测量结果。从图中可以明显地看到,在对测量向量进行归一化后,最小测量数只需要 120 次测量,而此时信号的稀疏度为 20。也就是说,这个最小测量数远比定理中要求的  $S^2 \lg N$  要小得多。由此可见,通过修改测量向量的形式从而减少观测次数是可

行的。

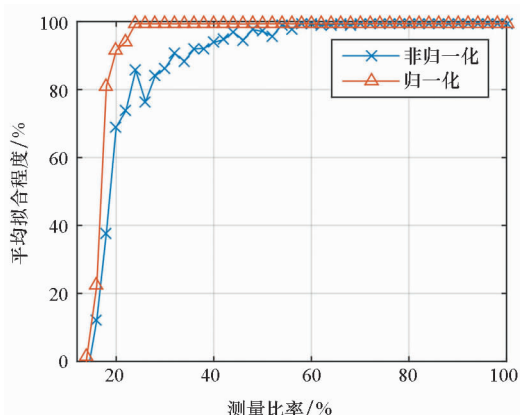


图 7 归一化对测量准确性的影响

Fig. 7 Influence of normalization on accuracy

## 4 结论

本文的重要贡献在于将压缩测量算法的思想从线性领域推广到非线性领域之中,即用测量的视角观察非线性系统的辨识问题。首先,仿真也证明了本文提出的测量方法在没有噪声的情况下是可以无损地获取系统模型,即使在有噪声的条件下,也可以保证稳定的测量结果。其次,在测量的视角下,辨识算法更关注于信号本身的信息量或者说信号的稀疏度,而对信号长度不敏感。这就给解决“维数灾难”问题指出了一条新的道路。随后,本文在仿真中详细分析了影响辨识结果的三个重要因素,即测量向量、测量数和测量噪声。特别是,本文提出了一种测量向量均值归零的方法,进一步缩减了最小测量数。这也就说明,通过设计合理的测量向量,是有可能进一步提高测量效率,使得测量数控制在稀疏度的常数倍范围内。在同等测量次数下,利用本文提出的方法可以获得更多的信息,即辨识模型的记忆长度和阶数规模都可以更大。当然,这两个参数越大,Volterra 模型可近似的非线性系统也就越复杂。

本文还简述了 Volterra 级数的物理含义。随后,本文将 Volterra 级数重新排列为符合压缩感知测量的形式,而且将原有测量矩阵重新排列为多个循环矩阵相连。这种方法不仅给 RIP 的证明提供了一个基础,还简化了辨识算法的程序设计。本文只是针对二阶以内的 Volterra 级数模型进行了仿真,不过,根据本文的方法可以推广到更高阶核的辨识上,该方向也将是未来的研究重点。此外,本文使用的是高斯分布组成的随机测量向量,随后将继续探索其他形式的测量向量,争取进一步地缩减最小测量数,提高测量效率。另外,本文

提出的方法不仅可以用于 Volterra 级数,还可以用于其他非线性核的线性组合表示的模型辨识问题之中。

## 参考文献 (References)

- [1] Yu C, Guan L, Zhu E N, et al. Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers[J]. *IEEE Transactions on Microwave Theory & Techniques*, 2012, 60(12): 4198–4208.
- [2] Shi C, Kajikawa Y. Identification of the parametric array loudspeaker with a volterra filter using the sparse NLMS algorithm [C]// *Proceedings of IEEE International Conference on Acoustics*, 2015: 3372–3376.
- [3] Turnip A, Tamba T A. Algorithmic modeling and analysis of nonlinear biological systems [C]// *Proceedings of the 10th Asian Control Conference*, 2015: 1–6.
- [4] Palm G, Poggio T. The Volterra representation and the wiener expansion: validity and pitfalls [J]. *SIAM Journal on Applied Mathematics*, 1977, 33(2): 195–216.
- [5] Feng G, Li H J, Dong J W, et al. Face recognition based on Volterra kernels direct discriminant analysis and effective feature classification[J]. *Information Sciences*, 2018, 441: 187–197.
- [6] Guo J, Yang S Y. Speaker recognition based on wavelet packet decomposition and Volterra adaptive model [C]// *Proceedings of 2nd IEEE International Conference on Computer and Communications*, 2017: 1952–1956.
- [7] Schölkopf B, Smola A J. *Learning with kernels: support vector machines, regularization, optimization, and beyond* [M]. *Publications of the American Statistical Association*, 2008.
- [8] Mathews V J, Sicuranza G L. *Polynomial signal processing*[M]. *John Wiley & Sons, Inc.*, 2000.
- [9] Birpoutsoukis G, Marconato A, Iataire J, et al. Regularized nonparametric Volterra kernel estimation [J]. *Automatica*, 2017, 82: 324–327.
- [10] Franz M O, Schölkopf B. A unifying view of wiener and Volterra theory and polynomial kernel regression [J]. *Neural Computation*, 2006, 18(12): 3097–3118.
- [11] Tan L. The application of compressed sensing in automotive radar signal processing for the target location [C]. *Chinese Control Conference*, 2016: 720–724.
- [12] Kalouptsidis N, Mileounis G, Babadi B, et al. Adaptive algorithms for sparse system identification [J]. *Signal Processing*, 2011, 91(8): 1910–1919.
- [13] Hans C. Bayesian lasso regression [J]. *Biometrika*, 2009, 96(4): 835–845.
- [14] Kekatos V, Angelosante D, Giannakis G B. Sparsity-aware estimation of nonlinear Volterra kernels [C]// *Proceedings of 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2010: 129–132.
- [15] Fan J Q, Li R Z. *Variable selection via nonconcave penalized likelihood and its oracle properties* [C]. *Taylor & Francis, Ltd. on behalf of the American Statistical Association*, 2001, 96(456): 1348–1360.
- [16] Syed M A, Mathews V J. QR-decomposition based algorithms for adaptive Volterra filtering [J]. *IEEE Transactions on Circuits & Systems I: Fundamental Theory & Applications*, 1993, 40(6): 372–382.
- [17] Donoho D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289–1306.
- [18] Candès E J. Compressive sampling [J]. *Proceedings of the International Congress of Mathematicians*, 2006: 1433–1452.
- [19] Bosworth B T, Stroud J R, Tran T D, et al. High-speed flow microscopy using compressed sensing with ultrafast laser pulses [J]. *Optics Express*, 2015, 23(8): 10521–32.
- [20] Ravishanker S, Bresler Y. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to MRI [J]. *Mathematics*, 2015, 2(3): 294–309.
- [21] Li L C, Li D J, Pan Z H. Compressed sensing application in interferometric synthetic aperture radar [J]. *Science China Information Sciences*, 2017, 60(10): 1–17.
- [22] Yin Y. The application of compressed sensing in automotive radar signal processing for the target location [C]. *Intelligent and Connected Vehicles Symposium*, 2017.
- [23] Kekatos V, Giannakis G B. Sparse Volterra and polynomial regression models: recoverability and estimation [J]. *IEEE Transactions on Signal Processing*, 2011, 59(12): 5907–5920.