

一种非白非高斯数据的数值仿真方法

刘旺锁, 王平波, 顾雪峰, 李大震

(海军工程大学, 武汉 430033)

摘要: 使用复合抽样法, 可以产生具有指定概率密度形式的加性分布非高斯序列。通过在极零图上直接指定数对极零点, 可以实现定性色化低阶自回归滤波器设计。把非高斯激励序列通过自回归滤波器, 即可得到非高斯信号处理仿真研究中频繁使用的非高斯有色序列。结合一组混合高斯有色数据数值仿真实例, 演示了这一由复合抽样法加定性色化构成的非白非高斯数据快捷数值仿真方法的有效性。

关键词: 非白; 非高斯; 复合抽样法; 自回归滤波; 混合高斯

中图分类号: TN911.7

文献标识码: A

文章编号: 1000-3630(2013)-03-0228-05

DOI 编码: 10.3969/j.issn1000-3630.2013.03.011

A simulation approach to colored non-Gaussian processes

LIU Wang-suo, WANG Ping-bo, GU Xue-feng, LI Da-zhen

(Naval University of Engineering, Wuhan 430033, China)

Abstract: By utilizing the composite sampling method, the non-Gaussian sequence with a certain additive distribution form can be generated. By assigning the couples of poles on the zero-pole plot, the low order autoregressive filter used for qualitative colorization can be realized. After driving the non-Gaussian sequence through the autoregressive filter, the colored non-Gaussian process, which is often used in the simulation study of non-Gaussian signal processing, is obtained. A numerical simulation instance of colored Gaussian mixture data is given to illustrate the performance of this approach of combining composite sampling and qualitative colorization.

Key words: non-white; non-Gaussian; composite sampling; autoregressive filter; Gaussian mixture

0 引言

在信号处理算法仿真研究中, 经常需要对非白非高斯数据进行数值仿真。根据设定的不同概率密度函数(Probability Density Function, PDF)和功率谱密度(Power Spectrum Density, PSD), 可以有不同的数值仿真方法。但无论选择怎样的方法, 一个最基本的评价标准就是, 所产生的数据的 PDF/PSD 一定要与设定足够逼近。否则, 可能会对后续处理和结果分析产生较大的负面影响。本文以混合高斯有色序列生成为例, 提出一种快速高效、简明直观的非白非高斯数据数值仿真方法。

与一般的高斯色数据产生方法^[1]类似, 这一非白非高斯数据产生方法的核心是自回归(Auto Regressive, AR)拟合指定功率谱技术^[2]: 将符合一定PDF的混合高斯激励序列, 通过一个具有指定PSD特性的AR滤波器, 即输出具备预设PDF和PSD

的混合高斯有色序列。这里的关键技术有两个: 其一是如何产生混合高斯白序列, 即如何进行混合高斯抽样; 其二是如何设计AR滤波器, 使其幅频响应符合指定的功率谱形状。

关于第一点, 有贝努利试验法^[3]、舍选抽样法^[4]、复合抽样法^[5]等。针对混合高斯分布属加分布的特性, 这里选用复合抽样法, 其意义简明直观, 效率也相对较高。关于第二点, 由于AR滤波器是线性滤波器, 故经典的高斯序列色化技术都可以应用于此, 包括从制订谱形出发的定量色化^[1]和从指定极点出发的定性色化^[6]。文献[6]详细探讨了这些非高斯抽样和色化方法, 但理论性较强, 讨论过于复杂, 本文从实用出发, 给出一种有效简便的、组合式的数据快捷生成方案。

1 混合高斯模型简介

一般地, 混合高斯(Gaussian Mixture, GM)PDF可表述为如式(1)所示的加分布形式。

$$f(u_n) = \sum_{m=1}^M \varepsilon_m \cdot f_m(u_n) \quad (1)$$

式中: u_n 为非高斯序列 $\mathbf{u}=[u_1, u_2, \dots, u_N]^T$ 的第 n 个样本值; M 为模型阶数; ε_m 为混合参数, 满足式(2)所

收稿日期: 2012-03-16; 修回日期: 2012-06-30

基金项目: 国家自然科学基金资助项目(51109218)

作者简介: 刘旺锁(1965—), 男, 江苏金坛人, 硕士生导师, 研究方向为声纳装备保障与效能评估、水声信号处理。

通讯作者: 刘旺锁, E-mail: blackberet@126.com

示的关系： f_m 为高斯分布 $N(\mu_m, \sigma_m^2)$ 的 PDF，如式(3)所示。

$$\sum_{m=1}^M \varepsilon_m = 1 \quad (2)$$

$$f_m(u_n) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(u_n - \mu_m)^2}{2\sigma_m^2}\right] \quad (3)$$

可以看到，GM 序列的 PDF 是一种纯粹的数学级数表达式，但它仍有自己明确的物理解释。事实上，从式(1)所示的 GM PDF 中，恰能看清混合高斯信号形成的物理机理，即混合高斯信号是由多个高斯源信号依据一定的概率分布组成的。具体而言， M 代表了组成非高斯信号的高斯源(亦称高斯分量)数量； f_m 就是第 m 个高斯源 G_m 的 PDF； ε_m 为当前样本 u_n 来自第 m 个高斯源的概率。显然，通过简单地调整 GM 参数 $\mathbf{g} = [\varepsilon_1, \dots, \varepsilon_M, \mu_1, \dots, \mu_M, \sigma_1^2, \dots, \sigma_M^2]^T$ 就可以拟合几乎任意的单钟型或多钟型 PDF。这就是 GM 模型在非高斯信号处理中倍受青睐的原因。

2 混合高斯抽样

应用蒙特卡罗方法的迭加原则，Marsaglia 提出了所谓的复合抽样法(Composition Sampling)^[2]。复合抽样法用于处理具有复合分布的随机变量抽样，而所谓复合分布就是指随机变量 x 服从的分布与另一个随机变量 y 有关的分布。加分布就是一种最简单的复合分布，对于它的随机变量抽样，应用复合抽样法将特别简便。而式(1)所示的 GM PDF 恰为一种加分布形式。

应用复合抽样法产生 GM 序列 $\{u_n\}$ 的具体步骤如下：

(1) 根据式(4)，由混合参数序列 $\{\varepsilon_m\}$ 构造一参考序列 $\{\eta_m\}$ ：

$$\eta_m = \sum_{i=1}^m \varepsilon_i \quad (4)$$

式中： $m \in [1, M]$ ，补充定义 $\eta_0 = 0$ 。不难发现， $\{\eta_m\}$ 为 $M+1$ 点升值序列。

(2) 生成相互独立的 M 组 N 点高斯源序列^[7] $\mathbf{x}^{(1)} \sim N(\mu_1, \sigma_1^2), \mathbf{x}^{(2)} \sim N(\mu_2, \sigma_2^2), \dots, \mathbf{x}^{(M)} \sim N(\mu_M, \sigma_M^2)$ ；

(3) 取 $(0,1)$ 上均匀分布的随机数 r_n ，解式(5)所示的不等式，得到高斯源编号 m ；

$$\eta_{m-1} < r_n \leq \eta_m \quad (5)$$

(4) 取第 m 个高斯源 $\mathbf{x}^{(m)}$ 的第 n 个样本 $x_n^{(m)}$ 为当前样本 u_n 。

(5) 从 $n=1$ 至 N 重复步骤(3)、(4)，即可得到 N 点 GM 序列。

由于阶数 M 一般不会很大，所以式(5)的求解

将很快捷，复合抽样法的效率将高于舍选抽样法。

对于二阶 GM 序列的生成，文献[1]中提出了一个很简便高效的方法——贝努利试验法。具体步骤如下：

(1) 生成相互独立的两组 N 点高斯序列 $\mathbf{x}^{(1)} \sim N(\mu_1, \sigma_1^2)$ 和 $\mathbf{x}^{(2)} \sim N(\mu_2, \sigma_2^2)$ ；

(2) 生成 N 点 0-1 二项分布序列 $\mathbf{r} \sim B(N, \varepsilon_1)$ ；

(3) 从 $n=1$ 到 N 抽样：若 $r_n=0$ ，则取 $x_n^{(1)}$ 为序列当前点 u_n ；若 $r_n=1$ ，则取 $x_n^{(2)}$ 为序列当前点 u_n 。

事实上，贝努利试验法的思路与复合抽样法一致。针对二阶情况，本文的复合抽样法可简化为如下所示的、与贝努利试验法类似的抽样方法：

(1) 生成相互独立的两组 N 点高斯序列 $\mathbf{x}^{(1)} \sim N(\mu_1, \sigma_1^2)$ 和 $\mathbf{x}^{(2)} \sim N(\mu_2, \sigma_2^2)$ ；

(2) 生成 N 点 $(0, 1)$ 上的均匀分布分布序列 $\mathbf{r} \sim U(0, 1)$ ；

(3) 从 $n=1$ 到 N 抽样：若 $r_n \leq \varepsilon_1$ ，则取 $x_n^{(1)}$ 为序列当前点 u_n ；否则，取 $x_n^{(2)}$ 为序列当前点 u_n 。

3 GM 序列的色化

不难发现，上述各方法产生的 GM 序列 $\{u_n\}$ ，各样本之间是相互独立的，因此其功率谱是白的。如果要得到混合高斯有色数据，还必须对其进行色化处理，而色化处理最常用的方法是进行 AR 滤波。要令滤波输出具有指定的功率谱形状，则必须恰当设计 AR 阶数 P 和系数 $\mathbf{a} = [a_1, a_2, \dots, a_P]^T$ ，此即 AR 拟合指定功率谱技术^[1,2,8]。如图 1 所示，根据需选择不同的功率谱模型，获得功率谱密度 Q ，对之进行逆傅里叶变换(IFFT)，得到相关函数 R ，解 Yule-Walker 方程，得到 LPC 参数：AR 系数 $\mathbf{a}$ 和激励方差 σ_u^2 ，由此建立起 AR 滤波器，然后用它对输入 GM 白序列 $\mathbf{u}$ 进行滤波，即可得到输出 GM 有色序列 $\mathbf{x}$ 。

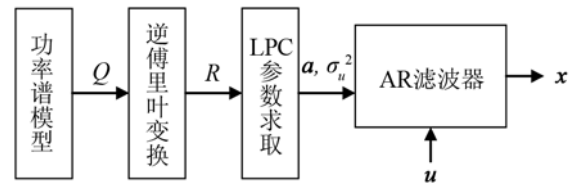


图1 AR 拟合指定功率谱技术
Fig.1 AR filter fitting assigned PSD

有时候，只要定性考察算法对有色混合高斯数据的处理效果，并不需要精确地设计数据的 PSD。比如，在仿真检验算法对窄带或宽带数据的处理能力时，仅要求产生的仿真数据具备窄带或宽带特性，而无需严格限定其 PSD 形状。对于这种应用情

形, 没有必要如图 1 所示那样进行复杂的全套 AR 拟合指定功率谱过程, 可以省却图中的前三个框, 直接设定 AR 参数 $\mathbf{a}$ 。不妨把这种色化称之为定性色化, 相应地, 把图 1 所示的色化称之为定量色化。

定性色化时所需的 AR 阶数不会太高, 一般为 2~4 阶。此时 AR 滤波器设计方法为: 在 AR 系统响应极零图上, 直接指定各极点, 进而获得合理的 AR 参数。下面即以 4 阶 AR(简记为 AR(4))为例演示这一过程。

AR(4)的系统函数如式(6)所示。

$$A(z) = \frac{\sigma_u^2}{1 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4} \quad (6)$$

式中: σ_u^2 为激励方差。

从极-零图上, 我们可以随意划定两组 (z_1 和 z_2 、 z_3 和 z_4) 单位圆内(确保系统稳定性)的极点, 如图 2 所示。

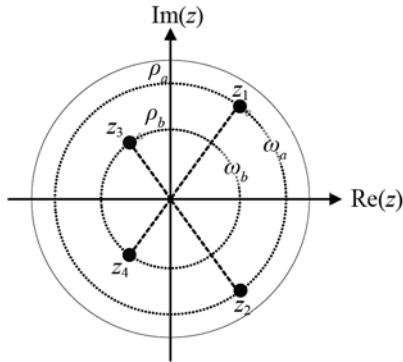


图 2 指定 AR(4)的极点

Fig.2 Appointment of AR(4)'s poles in narrow-band condition

图中: $\text{Re}(z)$ 、 $\text{Im}(z)$ 分别表示 z 的实部与虚部, 最外实线圆为单位圆; ρ_a 、 ρ_b 为两组极点的极径, 是 $(0,1]$ 之间的数; ω_a 、 ω_b 为两组极点的正角, 亦即圆周角, 是 $[0, \pi]$ 之间的数。这样, 两组极点分别指定为

$$\begin{aligned} z_{1,2} &= \rho_a \exp(\pm j\omega_a) \\ z_{3,4} &= \rho_b \exp(\pm j\omega_b) \end{aligned} \quad (7)$$

把这四个极点代入极点方程, 联立得到如式(8)所示的极点方程组, 解之即可得到 AR 参数 $\mathbf{a}=[a_1, a_2, a_3, a_4]^T$ 。

$$\begin{cases} 1 + a_1 z_1 + a_2 z_1^2 + a_3 z_1^3 + a_4 z_1^4 = 0 \\ 1 + a_1 z_2 + a_2 z_2^2 + a_3 z_2^3 + a_4 z_2^4 = 0 \\ 1 + a_1 z_3 + a_2 z_3^2 + a_3 z_3^3 + a_4 z_3^4 = 0 \\ 1 + a_1 z_4 + a_2 z_4^2 + a_3 z_4^3 + a_4 z_4^4 = 0 \end{cases} \quad (8)$$

按照图 2 中所示位置, $\rho_a=0.83$, $\rho_b=0.50$, $\omega_a=2\pi \cdot 0.15$, $\omega_b=2\pi \cdot 0.35$, 可解得 $\mathbf{a}=[-0.3879, 0.3654, 0.1610, 0.1722]^T$ 。于是可得到如图 3 所示的 PSD。这里设 $\sigma_u^2=34.80$ 。由图 3 可见, 这是一个窄

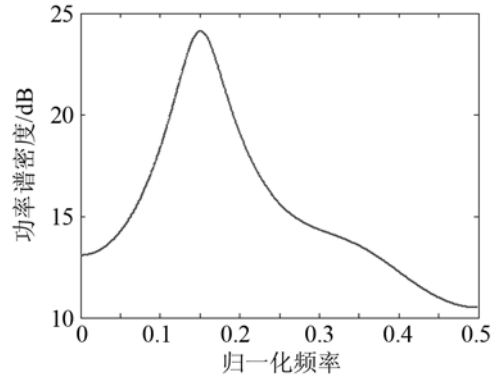


图 3 通过指定极点得到的 AR(4) PSD

Fig.3 Theoretical PSD of AR(4) in narrow-band condition

带情形。

通过调整极点位置, 可以得到所需的“调制”色谱。确定出 $\mathbf{a}$ 的值后, 再联合设定 σ_u^2 , 即可设计出符合需求的 AR 滤波器, 以之对白色 GM 激励序列滤波, 即可得到所需的有色 GM 序列。

4 仿真实例

下面以 3 阶 GM 有色序列生成为例演示本文所述方法的实际效果。

设定 GM 参数为 $\mathbf{g}=[\varepsilon, \mu, \sigma^2]^T$, $\varepsilon=[0.2, 0.5, 0.3]^T$, $\mu=[-15, 0, 15]^T$, $\sigma^2=[36, 25, 100]^T$ 。根据设定值绘出的理论 PDF 曲线如图 4 所示。

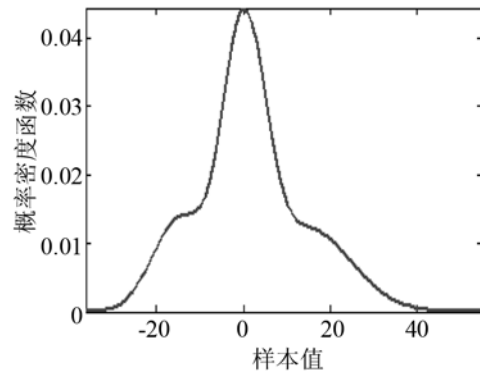


图 4 理论 PDF 曲线

Fig.4 Theoretical PDF for GM instance

使用复合抽样法产生采样点数 $N=10000$ 的白 GM 序列 $\mathbf{u}$, 波形如图 5 所示。图中仅绘出前 1000 点。由图可见, 波形中毛刺很多, 非高斯性表现较强。

为检验数据 $\mathbf{u}$ 是否具有指定的 PDF, 使用直方图统计其分布特性, 拟合为 PDF 曲线, 与理论 PDF 曲线相比较, 结果如图 6 所示。图中实线为理论设定值, 虚线为统计拟合值。可见, 两者吻合得很好, 说明所生成数据的确具有指定的概率密度特性。

接下来对 GM 序列 $\mathbf{u}$ 进行定性色化。第 3 节中已演示了窄带定性色化, 这里将引入宽带情形。同

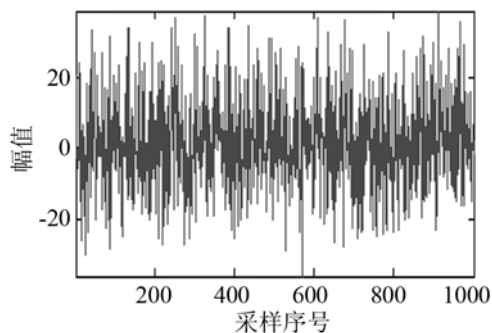


图5 仿真 GM 激励数据波形
Fig.5 Waveform of GM instance

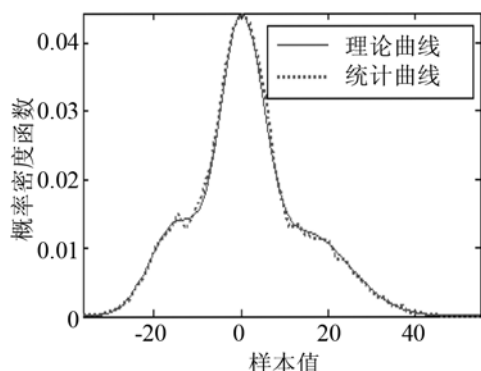


图6 仿真数据 PDF 与设定值比较(GM 实例)
Fig.6 Comparison between the theoretical and statistical PDFs for GM instance

样取 4 阶 AR 模型,其四个极点的指定如图 7 所示。这里, $\rho_a = \rho_b = 0.77$, $\omega_a = 2\pi \cdot 0.15$, $\omega_b = 2\pi \cdot 0.23$ 。

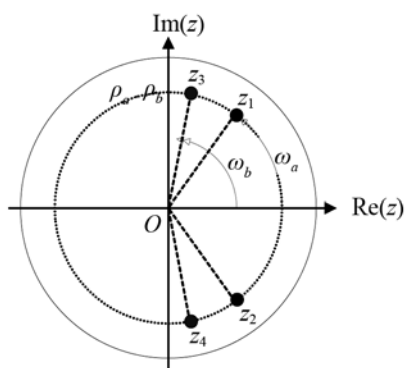


图7 仿真数据定性色化 AR(4)的极点
Fig.7 Appointment of AR(4)'s poles in wide-band condition

由图示极点位置,可解得 $\mathbf{a} = [-1.0982, 1.3605, -0.6511, 0.3515]^T$ 。同时,根据前面的 GM 参数设置,计算可得激励方差 $\sigma_u^2 = 156.86$ 。据此可得设定的理论 PSD 曲线如图 8 所示。

GM 激励序列 $\mathbf{u}$ 通过上述 AR(4)滤波器,得到非高斯有色序列输出 $\mathbf{x}$,其波形如图 9 所示。这里仍仅绘出前 1000 点。由图 9 可见,波形的非高斯性表现依然较强。

为检验数据 $\mathbf{x}$ 是否具有指定的 PSD,使用 Daniell 平滑周期图法统计其功率谱特性,拟合为

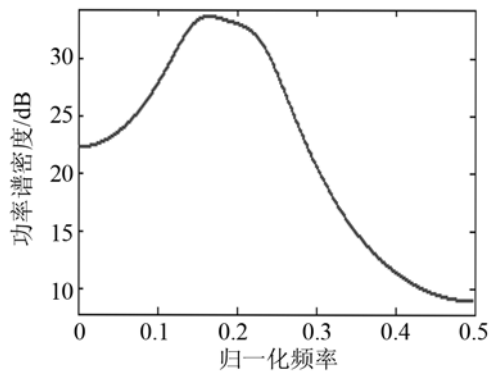


图8 理论 PSD 曲线
Fig.8 Theoretical PSD of AR(4) in wide-band condition

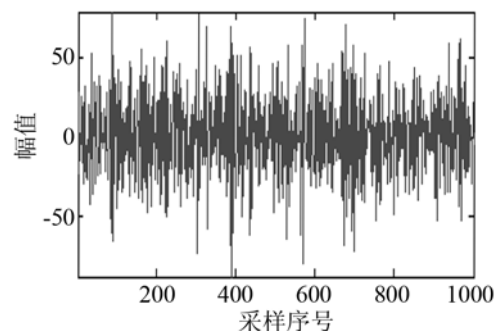


图9 仿真 GM 有色数据波形
Fig.9 Waveform of colored GM instance

PSD 曲线,与理论 PSD 曲线相比较,结果如图 10 所示。图中实线为理论设定值,虚线为统计拟合值。可以看到,两者吻合得很好,说明所生产数据的确具有指定的功率谱密度特性。

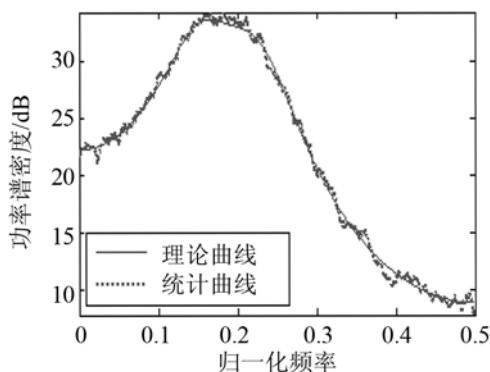


图10 仿真数据 PSD 与设定值比较(彩色 GM 实例)
Fig.10 Comparison between the theoretical and statistical PSDs for colored GM instance

综上,应用本文所述方法,可以产生具有指定 PDF 和 PSD 的有色非高斯数据。

5 结论

本文耦合复合抽样法和定性色化方法,给出了一种快捷的非高斯有色数据数值仿真方法。以混合高斯有色数据生成为例,演示了这一方法的有效

性。复合抽样法特别适合于具有加性分布形式的混合高斯数据抽样,在二阶情况下,抽样效率与贝努利试验法相当。而定性色化则可以快速地实现不严格要求谱形的窄带或宽带数据生成。

使用本文所述的方法,可以得到与指定 PDF、PSD 高度匹配的非高斯有色数据,基于这些数据进行的后续仿真信号处理能够真实地反映现实情况。

参 考 文 献

- [1] 王平波, 张明敏, 沈德刚. 三元组拖线阵声纳宽带最佳阵处理左右舷判别性能的仿真[J]. 信号处理, 2006, **22**(4): 554-558.
WANG Pingbo, ZHANG Mingmin, SHEN Degang. Simulation research of the port/starboard discrimination performance of hydrophone triplets towed line sonar by the optimal array processing algorithm in the case of broad band[J]. Signal Processing, 2006, **22**(4): 554-558.
- [2] 程广利, 王平波. AR 拟合指定功率谱技术功率对等问题研究 [J]. 声学技术, 2007, **26**(5): 123-126.
CHENG Guangli, WANG Pingbo. Research of the problem of power equivalence in the technique of AR fitting the power spectrum assigned [J]. Technical Acoustics, 2007, **26**(5): 123-126.
- [3] Debasis Sengupta, Steven Kay. Efficient estimation of parameters for non-Gaussian autoregressive processes[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989, **37**(6): 785-794.
- [4] 马文淦. 计算物理学[M]. 北京: 科学出版社, 2005: 24-55.
MA Wengan. Computational physics[M]. Beijing: Science Publishing House, 2005: 24-55.
- [5] 蒋瑜, 陶俊勇, 王得志, 等. 一种新的非高斯随机振动数值模拟方法[J]. 振动与冲击, 2012, **31**(19): 169-173.
JIANG Yu, TAO Junyong, WANG Dezhi. A novel approach for numerical simulation of a non-Gaussian random vibration[J]. Journal of Vibration and Shock, 2012, **31**(19): 169-173.
- [6] 王平波, 蔡志明, 姜可宇. 混合高斯有色数据的生成方法[J]. 声学与电子工程, 2007, **22** (1): 7-11.
WANG Pingbo, CAI Zhiming, JIANG Keyu. The generating method of colored non-Gaussian processes[J]. Acoustics and Electronics Engineering, 2007, **22**(1): 7-11.
- [7] 蔡坤宝, 王成良, 陈曾汉. 产生标准高斯白噪声序列的方法[J]. 中国电机工程学报, 2004, **24**(12): 207-211.
CAI Kunbao, WANG Chenliang, CHEN Zenghan. A method for generating standard Gaussian white noise sequences[J]. Proceedings of the CSEE, 2004, **24**(12): 207-211.
- [8] 郑兆宁, 向大威. 水声信号被动检测与参数估计理论[M]. 北京: 科学出版社, 1983: 685.
ZHENG Zhaoning, XIANG Dawei. Theory of the passive detection and estimation for underwater acoustical signal[M]. Beijing: Science Publishing House, 1983: 685.