Generating Random Variates for
Stable Sub-Gaussian Processes with Memory

A. Mahmood[a,*], M. Chitre[a,b]

[a]*Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, Singapore 119227*
[b]*Department of Electrical and Computer Engineering, Block E4, 4 Engineering Drive 3, Singapore 117583*

Abstract

We present a computationally efficient method to generate random variables from a univariate conditional probability density function (PDF) derived from a multivariate α-sub-Gaussian (αSG) distribution. The approach may be used to sequentially generate variates for sliding-window models that constrain immediately adjacent samples to be αSG random vectors. We initially derive and establish various properties of the conditional PDF and show it to be equivalent to a Student's t-distribution in an asymptotic sense. As the αSG PDF does not exist in closed form, we use these insights to develop a method based on the rejection sampling (accept-reject) algorithm that allows generating random variates with computational ease.

Keywords: random number generation, symmetric α-stable distribution, α-sub-Gaussian distribution, Student's t-distribution, rejection sampling.

1. Introduction

The heavy-tailed *stable* distribution has been extensively used in the literature to model impulsive data [11, 6, 15]. Heavy-tailed distributions assign non-negligible probabilities to outliers and therefore offer good fits to amplitude distributions of impulsive datasets [11]. The motivation for employing stable models stems from the generalized central limit theorem (GCLT) [21], which states that the sum of independent and identically distributed (IID) random variables (vectors) converges to a *stable* distribution as the number of elements in the sum approaches infinity [21, 15]. The GCLT is in fact the central limit theorem (CLT) but with the power constraint removed. The last statement essentially implies that the well-known Gaussian distribution is part of the stable family. Moreover, it is the only member to have light (exponential) tails [21, 15]. As the validity of Gaussian models is primarily attributed to the CLT [17], the GCLT offers a similar argument for

heavy-tailed stable models when the process is impulsive in nature [15, 11]. From an engineering perspective, there are a number of practical scenarios where the ambient noise process is known to be impulsive. The warm shallow underwater channel, powerlines and interference-prone wireless networks are a few examples where such noise is prevalent [6, 23].

In the literature, impulsive noise processes are typically assumed to be white, i.e., the samples are symmetric IID random variables [6, 5, 2]. Though such assumptions offer mathematical tractability in terms of developing optimized routines, they are far from realistic [3, 10, 7, 4]. In practice, noise is seldom white and therefore has memory. Thus, at any given time, the current noise sample depends on a number of previous samples. To incorporate this dependence within the noise framework, several models have been proposed in the literature [3, 10, 7, 6]. Recently, the *stationary α-sub-Gaussian noise with memory order m (αSGN(m))* process was proposed to model the ambient noise in warm shallow underwater environments [7]. As highlighted by its name, the αSGN(m) model is based on the α-sub-Gaussian (αSG) distribution, which is a subclass of the stable family with the

---

*Corresponding author
Email addresses: `tmsahme@nus.edu.sg` (A. Mahmood), `mandar@nus.edu.sg` (M. Chitre)

added characteristic of being symmetric and elliptical as well [14]. The new model is particularly adept in tracking the dependence between samples of the noise process whilst constraining the amplitude distribution to be that of a symmetric α-stable (SαS) random variable and is shown to outperform contemporary colored and white models in this regard.

Generating random variates is an important component of simulation-based performance analysis of systems, schemes and algorithms. The αSGN(m) model is based on a sliding-window type framework and constrains samples within the window to be αSG [7]. Therefore, a noise sample is returned from a *univariate conditional distribution* of a *multivariate αSG* distribution. As shown later, this is a computationally demanding task, especially when a large number of variates are required. However, by expressing the conditional density in a suitable form and taking advantage of its properties, it is possible to compute realizations in a less time-consuming manner.

The primary contribution of this paper is to offer a way to *efficiently* generate realizations for αSG processes that are based on a sliding-window framework. The αSGN(m) model falls within this category. Generating *independent* outcomes from multivariate αSG distributions is a straightforward task and may be accomplished with computational ease [14, 9]. However, extending this efficiency to processes with memory is not trivial. To accomplish this, we investigate the properties of a univariate conditional distribution derived from a (general) multivariate αSG probability density function (PDF). We show that this converges to a Student's t-distribution in an asymptotic sense and express the latter's parameters in terms of the formers. These properties are then exploited to find suitable majorizing functions which are used to efficiently generate random variates by employing the established rejection sampling (accept-reject) method [19]. Further still, the general αSG PDF cannot be expressed in closed-form [14]. For every instance of its argument, the PDF is evaluated by numerically integrating over a *heavy-tailed function* [11, 14]. The same holds true for its corresponding conditional PDFs. We investigate the underlying function and show that it can be evaluated by a one-time tabulation and interpolation over a certain range, after which limiting expressions may be applied. Employing this routine in conjunction with the optimized setting of the re-

jection sampling algorithm substantially reduces the time taken for generating the realizations. It takes approximately a second to sequentially generate 10,000 samples of αSGN(m) on a 3.70GHz processor.

This paper is organized as follows: In Section 2, we briefly summarize stable, SαS and multivariate αSG distributions. We then derive the univariate conditional density for the latter and comment on its properties in Section 3. Using these insights, we discuss random number generation from the conditional PDF via rejection sampling in Section 4. We wrap up by presenting the conclusions in Section 5.

2. Summary of Concepts & Notation

2.1. Stable Distributions

A *random vector* $\vec{X} \in \mathbb{R}^d$ is stable if and only if

$$a_1 \vec{X}^{(1)} + a_2 \vec{X}^{(2)} \stackrel{d}{=} b\vec{X} + c, \qquad (1)$$

where $a_1, a_2, b, c \in \mathbb{R}$, $\vec{X}^{(i)}$ are IID copies of $\vec{X}$ and $\stackrel{d}{=}$ implies equality in distribution [15, 21]. One of the implications of (1) is that the distribution of $\vec{X}$ is *conserved* under linear transformations up to location and scale. This is termed as the stability property and is unique to this family of distributions [11, 21]. From (1), one notes that if $\vec{X}$ is stable, then any permutation of $\vec{X}$ is stable as well. Further still, if $\vec{X} = [\vec{X}_1^\mathsf{T}, \ \vec{X}_2^\mathsf{T}]^\mathsf{T}$ such that $\vec{X}_1 \in \mathbb{R}^m$ and $\vec{X}_2 \in \mathbb{R}^{d-m}$, then the marginal random vectors $\vec{X}_1$ and $\vec{X}_2$ are stable too [21].

A *univariate* stable distribution is parameterized by four parameters, namely the characteristic exponent $\alpha \in (0, 2]$, the skew parameter $\beta \in [-1, +1]$, the scale $\delta \in (0, +\infty)$ and the location $\mu \in (-\infty, +\infty)$ [15, 21]. In notational form, it is denoted by $\mathcal{S}(\alpha, \beta, \delta, \mu)$ [11]. Stable distributions are generally heavy-tailed, with the heaviness solely determined by α. As the latter increases, the tails become *increasingly lighter*, ultimately converging to a Gaussian distribution with mean μ and variance $2\delta^2$, i.e., $\mathcal{N}(\mu, 2\delta^2)$ for $\alpha = 2$ [15]. If $\vec{X}$ is a stable random vector, then its elements are stable random variables and have the same α.

A *univariate* SαS distribution is stable, but with β and μ set to zero [11, 21]. Consequently, its PDF symmetric about zero for any α. We denote such distributions by $\mathcal{S}(\alpha, \delta)$. For $\alpha = 2$, $\mathcal{S}(2, \delta) \stackrel{d}{=}$

$\mathcal{N}(0, 2\delta^2)$. Extending this to the multidimensional case, we note that $\vec{X} \in \mathbb{R}^d$ is SαS if it satisfies (1) and its PDF $f_{\vec{X}}(\mathbf{x})$ is symmetric, i.e.,

$$f_{\vec{X}}(\mathbf{x}) = f_{\vec{X}}(-\mathbf{x}), \qquad (2)$$

where $\mathbf{x} \in \mathbb{R}^d$ is a sample outcome of $\vec{X}$. Moreover, all marginal distributions of $\vec{X}$ are SαS and each element of $\vec{X}$ is SαS with the same α [21, 11].

One disadvantage of employing stable random variables is the general lack of closed-form PDFs. Exceptions are the Gaussian ($\alpha = 2$) case and the Cauchy case ($\alpha = 1$). Evaluating the PDF at a point requires calling a numerical routine, which is computationally taxing [11, 21]. This is more prominent in multivariate cases. However, in some instances, certain conducive properties may be exploited to reduce evaluation times. The αSG distribution is such an example and is discussed next.

2.2. The αSG Distribution

An αSG distribution is *heavy-tailed* SαS with the added constraint of being *elliptical* as well [14, 21]. More precisely, the random vector $\vec{X} \in \mathbb{R}^d$ is αSG ($\alpha \neq 2$) if it can be expressed as

$$\vec{X} = \sqrt{A}\vec{G}, \qquad (3)$$

where $A \sim \mathcal{S}(\frac{\alpha}{2}, 1, 2(\cos(\frac{\pi\alpha}{4}))^{2/\alpha}, 0)$ and $\vec{G} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is a d-dimensional Gaussian random vector with the all-zero location vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ [14, 21, 11]. Moreover, A and $\vec{G}$ are independent. The matrix $\boldsymbol{\Sigma}$ is also called the *shape matrix* of $\vec{X}$ [14]. The *marginal distribution* corresponding to any tuple of elements in $\vec{X}$ is also an αSG distribution and is therefore elliptic as well [14]. Denoting the ith diagonal element of $\boldsymbol{\Sigma}$ as σ_{ii}^2, the distribution of the ith element in $\vec{X}$ is $\mathcal{S}(\alpha, \sigma_{ii})$.

Before we present the PDF of $\vec{X}$, it is pertinent to define the *standard isotropic* αSG vector

$$\begin{aligned} \vec{Y} &= \boldsymbol{\Sigma}^{-1/2}\vec{X} \\ &= \sqrt{A}\vec{G}_Y, \end{aligned} \qquad (4)$$

where $\vec{G}_Y = \boldsymbol{\Sigma}^{-1/2}\vec{G} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $\mathbf{I}_d$ is the $d \times d$ identity matrix. The lower-triangular matrix $\boldsymbol{\Sigma}^{1/2}$ arises from the Cholesky decomposition of $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})^\mathsf{T}$, as the latter is a symmetric positive semi-definite matrix [22, 18]. In turn, the PDF

of $\vec{Y}$ can be written in terms of the PDF of the *radial* random variable $R = \|\vec{Y}\| = (\vec{Y}^\mathsf{T}\vec{Y})^{1/2}$ [14]. The latter can be expressed as

$$f_R(r; \alpha, d) = 2r \int_0^\infty f_A(r^2/t; \alpha) \frac{f_{\chi^2}(t; d)}{t} \mathrm{d}t, \quad (5)$$

where $f_A(t; \alpha)$ is the PDF of A (with argument t, parameterized by α) and $f_{\chi^2}(t; d)$ is a χ^2 PDF with d degrees of freedom. The PDF of $\vec{Y}$ is then given by

$$f_{\vec{Y}}(\mathbf{y}; \alpha, d) = \frac{\Gamma(d/2)}{2\pi^{d/2}}\|\mathbf{y}\|^{1-d}f_R(\|\mathbf{y}\|; \alpha, d). \quad (6)$$

By using the expression

$$v(r; \alpha, d) = |r|^{1-d}f_R(r; \alpha, d), \qquad (7)$$

we can express (6) in a more simplified form:

$$f_{\vec{Y}}(\mathbf{y}; \alpha, d) = \frac{\Gamma(d/2)}{2\pi^{d/2}}v(\|\mathbf{y}\|; \alpha, d). \qquad (8)$$

As $\vec{X} = \boldsymbol{\Sigma}^{1/2}\vec{Y}$, one can invoke simple matrix transformations to get

$$\begin{aligned} f_{\vec{X}}(\mathbf{x}; \alpha, \boldsymbol{\Sigma}, d) &= \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}}f_{\vec{Y}}(\boldsymbol{\Sigma}^{-1/2}\mathbf{x}; \alpha, d) \\ &= \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}}\frac{\Gamma(d/2)}{2\pi^{d/2}}v(\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|; \alpha, d). \end{aligned} \qquad (9)$$

Like most SαS distributions, the PDF in (9) cannot be written in closed-form [14]. The only exception to this is the Cauchy case ($\alpha = 1$) [14, 20], for which we have

$$f_{\vec{X}}(\mathbf{x}; 1, \boldsymbol{\Sigma}, d) = \\ \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}}\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{1}{2})\pi^{d/2}(1 + \|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|^2)^{\frac{d+1}{2}}}. \quad (10)$$

As $f_{\vec{X}}(\mathbf{x}; \alpha, \boldsymbol{\Sigma}, d)$ depends on $v(r; \alpha, d)$, one can express (9) in terms of a *single integral*, irrespective of d [14]. Consequently, the added computational cost of calculating higher dimensional PDFs comes primarily from the evaluating the determinant in (9). In Figs. 1 and 2, we plot $v(r; \alpha, d)$ for $\alpha = 1.5$ and $\alpha = 1.8$, respectively, for $d \in \{1, 2, \ldots, 10\}$ on a log-log scale. One notes that the curves are a non-increasing function of r. Also, there are two *linear* regions in each curve: a *flat* region as $r \to 0$ and
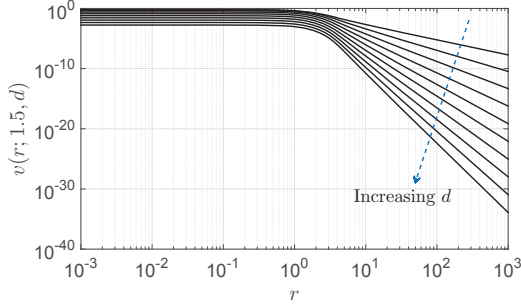
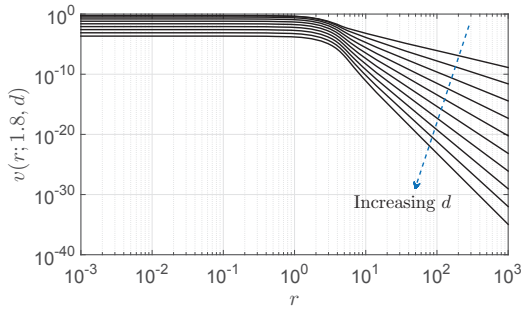Figure 1: $v(r; 1.5, d)$ for $d \in \{1, \ldots, 10\}$.

Figure 2: $v(r; 1.8, d)$ for $d \in \{1, \ldots, 10\}$.

a *descending* region for large r. These characteristics are observed for other values of α and d. In fact, from the series representation of $f_R(r)$ in [14], $v(r; \alpha, d)$ has the limiting properties

$$\lim_{r \to \infty} r^{\alpha+d} v(r; \alpha, d) = \alpha k_1(\alpha, d) \quad \text{and} \quad (11)$$

$$\lim_{r \to 0} v(r; \alpha, d) = k_2(\alpha, d), \quad (12)$$

where

$$k_1(\alpha, d) = 2^\alpha \frac{\sin(\frac{\pi\alpha}{2})\Gamma(\frac{\alpha+2}{2})\Gamma(\frac{\alpha+d}{2})}{\frac{\pi\alpha}{2}\Gamma(\frac{d}{2})} \quad \text{and}$$

$$k_2(\alpha, d) = \frac{4\Gamma(\frac{d}{\alpha})}{\alpha 2^d \Gamma(\frac{d}{2})^2}.$$

Both (11) and (12) offer close approximations for the downwards sloping and flat (near-zero) regions in $v(r; \alpha, d)$, respectively. We use these observations to our advantage later on.

Remark 1. To get $v(r; \alpha, d)$, we first evaluate $f_R(r; \alpha, d)$ via (5) and then invoke (7). For general α, $f_R(r; \alpha, d)$ cannot be expressed in closed form

and needs to be evaluated numerically [14]. Moreover, it depends on $f_A(t; \alpha)$ which also requires numerical integration [12]. We employ the STABLE package [13, 16] to get $f_A(t; \alpha)$ and employ numerical integration methods in MATLAB [8] to evaluate (5). Thereafter, $v(r; \alpha, d)$ is calculated from (7).

3. The Conditional αSG Distribution

We now derive and investigate the properties of the conditional PDF of a single component in an αSG random vector. The results in this section offer us the necessary tools to develop the algorithms proposed in Section 4.

3.1. Derivation & Basic Properties

If $\vec{X} \in \mathbb{R}^d$ is an αSG random vector, we can express it as $\vec{X} = [\vec{X}_1^\mathsf{T}, X_2]^\mathsf{T}$, where $\vec{X}_1 \in \mathbb{R}^{d-1}$ and $X_2 \in \mathbb{R}$. Similarly, as the underlying $\boldsymbol{\Sigma}$ is a symmetric matrix, it can be expressed in the block form

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^\mathsf{T} & \sigma_{22}^2 \end{bmatrix}, \quad (13)$$

where $\boldsymbol{\Sigma}_{11}$ is of size $(d-1) \times (d-1)$, $\boldsymbol{\Sigma}_{12}$ is $(d-1) \times 1$ and σ_{22}^2 is a scalar. Note that $\boldsymbol{\Sigma}_{11}$ is a real covariance matrix, and is therefore symmetric positive semi-definite [17]. From the discussion in Section 2.2, $\vec{X}_1$ is αSG with underlying covariance matrix $\boldsymbol{\Sigma}_{11}$. Therefore the PDF of $X_2|\vec{X}_1 = \mathbf{x}_1$ is

$$f_{X_2|\vec{X}_1}(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) = \frac{f_{\vec{X}}(\mathbf{x}; \alpha, \boldsymbol{\Sigma}, d)}{f_{\vec{X}}(\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}_{11}, d-1)}, \quad (14)$$

where $\mathbf{x} = [\mathbf{x}_1^\mathsf{T}, x_2]^\mathsf{T}$. For brevity, we employ the notation $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ to represent $f_{X_2|\vec{X}_1}(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$. On substituting (9) in (14), we get

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) = \sqrt{\frac{\det \boldsymbol{\Sigma}_{11}}{\det \boldsymbol{\Sigma}}} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})\sqrt{\pi}}$$
$$\times \frac{v(\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|; \alpha, d)}{v(\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d-1)}. \quad (15)$$

It is hard to discern any properties from (15). We therefore express it in a more interpretive form.

4

Using block matrix manipulations, one can write

$$\boldsymbol{\Sigma}^{1/2} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{1/2} & \mathbf{0} \\ \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1/2} & \sqrt{\kappa} \end{bmatrix}, \qquad (16)$$

where $\kappa = \sigma_{22}^2 - \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$ is the Schur's complement of $\boldsymbol{\Sigma}_{11}$ [18, pg. 47]. Details are provided in Appendix A. Using inversion identities for block matrices [18, pg. 47], we have

$$\boldsymbol{\Sigma}^{-1/2} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1/2} & \mathbf{0} \\ -\frac{\boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1}}{\sqrt{\kappa}} & 1/\sqrt{\kappa} \end{bmatrix}, \qquad (17)$$

and therefore

$$\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Sigma}^{-1/2})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1/2}$$
$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \frac{\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1}}{\kappa} & -\frac{\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}}{\kappa} \\ -\frac{\boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1}}{\kappa} & \frac{1}{\kappa} \end{bmatrix}. \qquad (18)$$

The latter follows as $\boldsymbol{\Sigma}_{11}^{-1} = (\boldsymbol{\Sigma}_{11}^{-1})^{\mathsf{T}}$. From (18) and the fact that $\mathbf{x} = [\mathbf{x_1}^{\mathsf{T}}, x_2]^{\mathsf{T}}$, we have

$$\|\boldsymbol{\Sigma}^{-1/2} \mathbf{x}\|^2 = \mathbf{x}^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \mathbf{x}$$
$$= \|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1\|^2 + \frac{(x_2 - \hat{\mu})^2}{\kappa}, \qquad (19)$$

where $\hat{\mu} = \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{x}_1$. Also, as κ is a scalar, we may use the identity $\det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}_{11} \det \kappa$ [18, pg. 46] to get

$$\kappa = \frac{\det \boldsymbol{\Sigma}}{\det \boldsymbol{\Sigma}_{11}}. \qquad (20)$$

As $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{11}$ are positive semi-definite matrices, $\kappa \in \mathbb{R}^+$. On substituting (19) and (20) back in (15), we finally get

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) = \frac{1}{\sqrt{\kappa \pi}} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})}$$
$$\times \frac{v\left(\left(\|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1\|^2 + \frac{(x_2 - \hat{\mu})^2}{\kappa}\right)^{1/2}; \alpha, d\right)}{v(\|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1\|; \alpha, d-1)}. \qquad (21)$$

For the special case of $\vec{X}$ being a standard isotropic random vector, (21) reduces to the simplified form:

$$f(x_2|\mathbf{x}_1; \alpha, \mathbf{I}_d, d) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})}$$

$$\times \frac{v((\|\mathbf{x}_1\|^2 + x_2^2)^{1/2}; \alpha, d)}{v(\|\mathbf{x}_1\|; \alpha, d-1)}. \qquad (22)$$

Therefore, from (21) and (22), we have the relationship:

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) = \frac{1}{\sqrt{\kappa}} f\left(\frac{x_2 - \hat{\mu}}{\sqrt{\kappa}} | \hat{\mathbf{x}}_1; \alpha, \mathbf{I}_d, d\right), \qquad (23)$$

where $\hat{\mathbf{x}}_1 = \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1$.

Though cumbersome, the form in (21) offers us some clear deductions:

1. The conditional PDF is symmetric about $\hat{\mu}$.

2. From (19), for a given $\|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1\|$, one notes that $\|\boldsymbol{\Sigma}^{-1/2} \mathbf{x}\|$ is a monotonically increasing function of $|x_2 - \hat{\mu}|$. This result, along with the fact that $v(r; \alpha, d)$ is a non-increasing function of r, constrains (21) to be unimodal. Due to the latter's symmetry, the mode is located at $\hat{\mu}$.

3. For any $\|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_1\| \in \mathbb{R}^+$, the scale of $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ is proportional to $\sqrt{\kappa}$. This is more apparent in (23).

Due to the relationship in (3), it is not surprising that (21) shares the above properties with the corresponding conditional distribution of the underlying $\vec{G}$. More explicitly, if $\vec{G} = [\vec{G}_1^{\mathsf{T}}, G_2]^{\mathsf{T}}$, then the distribution of $G_2|\vec{G}_1 = \mathbf{g}_1$ follows $\mathcal{N}(\boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{g}_1, \kappa)$. The properties listed above extend to the Gaussian case, but with $\mathbf{x}_1$ replaced by $\mathbf{g}_1$. Do note that the conditional Gaussian PDF is a closed-form expression. On the other hand, (21) cannot be expressed in closed-form for general α. The only exception to this is the Cauchy case, which we briefly discuss next.

3.2. The Cauchy Case

It is well known that the multivariate Cauchy distribution is also a multivariate Student's t-distribution [20]. For the latter, the corresponding conditional distributions are also Student's t-distributions [20]. Therefore, the univariate conditional PDF of a multivariate Cauchy distribution will be Student's t as well. More explicitly, from

5

(10) and (14), we have

$$f(x_2|\mathbf{x}_1; 1, \boldsymbol{\Sigma}, d) = \sqrt{\frac{\det \boldsymbol{\Sigma}_{11}}{\det \boldsymbol{\Sigma}}} \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})\sqrt{\pi}}$$
$$\times \frac{(1 + \|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|^2)^{-(d+1)/2}}{(1 + \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2)^{-d/2}}. \quad (24)$$

On substituting (19) and (20) in the above expression and simplifying, we have

$$f(x_2|\mathbf{x}_1; 1, \boldsymbol{\Sigma}, d) = \frac{1}{\sqrt{\kappa\pi}} \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}$$
$$\times \frac{\left(1 + \frac{(x_2-\hat{\mu})^2}{\kappa(1+\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2)}\right)^{-(d+1)/2}}{\sqrt{1 + \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2}}. \quad (25)$$

The *univariate* Student's t-PDF is given by

$$f_S(s; \nu, \delta, \mu) = \frac{\Gamma(\frac{\nu+1}{2})}{\delta\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}$$
$$\times \left(1 + \frac{(s-\mu)^2}{\nu\delta^2}\right)^{-(\nu+1)/2}, \quad (26)$$

where $\nu \in \mathbb{R}^+$ is the degrees of freedom, $\delta \in \mathbb{R}^+$ is the scale of the distribution and $\mu \in \mathbb{R}$ is the location parameter [20]. We denote it by $\mathcal{T}(\nu, \delta, \mu)$ and note it to be unimodal and symmetric about μ. Like the SαS family, the tails of a Student's t-distribution are algebraic, the heaviness of which is controlled by ν. On comparing (25) with (26), one notes that

$$f(x_2|\mathbf{x}_1; 1, \boldsymbol{\Sigma}, d)$$
$$= f_S\left(x_2; d, \sqrt{\frac{\kappa(1 + \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2)}{d}}, \hat{\mu}\right). \quad (27)$$

The equality in (27) is a known result [20], but it raises an important question: Is the Student's t-distribution related to the conditional αSG PDF in (21) for any $\alpha \neq 1$ and $d \in \mathbb{Z}^+$? We answer this next.

3.3. Convergence to the Student's t-distribution

The PDF in (21) converges to that of a univariate Student's t only in a limiting sense. This is highlighted by the following proposition:

Proposition 1. *If $\vec{X} \in \mathbb{R}^d$ is an αSG random vector,* *then for $0 < \alpha < 2$, $\hat{\nu} = \alpha + d - 1$ and $\hat{\mu} = \boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1$,*

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$$
$$\sim f_S\left(x_2; \hat{\nu}, \sqrt{\frac{\kappa}{\hat{\nu}}}\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|, \hat{\mu}\right) \quad (28)$$

$$\forall\, x_2 \in \mathbb{R}, \text{ as } \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \to \infty.$$

PROOF. From (19), we have $\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\| \geq \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|$. Therefore, as $\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \to \infty$, then $\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\| \to \infty$ as well. We can thus substitute (11) for both the numerator and denominator terms in (21) to get

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) \sim \frac{1}{\sqrt{\kappa\pi}} \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})} \frac{k_1(\alpha, d)}{k_1(\alpha, d-1)}$$
$$\times \frac{\left(\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2 + \frac{(x_2-\boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1)^2}{\kappa}\right)^{-(\alpha+d)/2}}{\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^{-(\alpha+d-1)}} \quad (29)$$

as $\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \to \infty$, where $\sim$ denotes *asymptotic equivalence*. On simplifying the above expression, we get

$$f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) \sim \frac{1}{\sqrt{\kappa\pi}} \frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(\frac{\alpha+d-1}{2})}$$
$$\times \frac{\left(1 + \frac{(x_2-\boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1)^2}{\kappa\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2}\right)^{-(\alpha+d)/2}}{\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|}. \quad (30)$$

On comparison with (26), we deduce that the right-hand side of (30) is equivalent to $f_S(x_2; \hat{\nu}, \sqrt{\frac{\kappa}{\hat{\nu}}}\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|, \hat{\mu})$, where $\hat{\nu} = \alpha + d - 1$ and $\hat{\mu} = \boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1$. $\square$

Remark 2. In (29), the asymptotic expression in (11) is substituted for both $v(\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\|; \alpha, d)$ and $v(\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d-1)$ to get (28). In the latter case, $\sim$ denotes asymptotic equivalence, i.e., $\lim f(x_2|\ldots)/f_S(x_2;\ldots) = 1$ as $\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \to \infty$. We note that (11) *approximates* the downward linear sloping region of $v(r; \alpha, d)$ very well (revert to the example curves in Figs. 1 & 2). Therefore, as $\|\boldsymbol{\Sigma}^{-1/2}\mathbf{x}\| \geq \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|$, one would expect $f_S(x_2;\ldots)$ in (28) to be a good *approximation* for $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ when $\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|$ is large. For a given $\boldsymbol{\Sigma}$, the latter arises when $\|\mathbf{x}_1\|$ is large. Therefore, if *at least one* element in $\mathbf{x}_1$ is impul-

sive, $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ is approximated well by (28). Specifically, for the Cauchy case, one can see that (28) (with $\alpha = 1$) offers a good approximation to (27) when $\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \gg 1$. On a different note, we see that both $\hat{\mu}$ and the scale in (28) diverge as $\|\mathbf{x}_1\| \to \infty$. One may be tempted to suggest that in such a scenario (28) may not be influenced by α or d. However, we see that the tails of the PDF depend on $\hat{v}$, which is independent of $\mathbf{x}_1$. Also, when $\|\mathbf{x}_1\|$ is finite but large, the asymptotic approximation applies. This results in both $\hat{\mu}$ and $\sqrt{\kappa/\hat{\nu}}\|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|$ being finite.

Corollary 1. *If $\vec{X} \in \mathbb{R}^d$ is a standard isotropic αSG random vector, then for $0 < \alpha < 2$ and $\hat{\nu} = \alpha + d - 1$,*

$$f(x_2|\mathbf{x}_1; \alpha, \mathbf{I}_d, d) \sim f_S\left(x_2; \hat{\nu}, \|\mathbf{x}_1\|/\sqrt{\hat{\nu}}, 0\right) \quad (31)$$

$$\forall\, x_2 \in \mathbb{R}, \text{ as } \|\mathbf{x}_1\| \to \infty.$$

PROOF. For the standard isotropic case, we have $\boldsymbol{\Sigma} = \mathbf{I}_d$. From the latter's block form in (13), we have $\boldsymbol{\Sigma}_{11} = \mathbf{I}_{d-1}$ and $\hat{\mu} = \boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1 = 0$. Also, from (20), we have $\kappa = 1$. These are substituted in (28) to get (31). $\qquad\square$

4. Generating Random Variates

Now that we have established various properties of the conditional αSG distribution, we will address the primary objective of this paper, i.e., develop a procedure to generate random variates for models that employ such distributions. We start off by briefly discussing how this can be accomplished for the underlying Gaussian distribution and the Cauchy case.

As discussed in Section 3.1, $G_2|\vec{G}_1 = \mathbf{g}_1 \sim \mathcal{N}(\boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{g}_1, \kappa)$. Therefore, the corresponding random variable can be expressed as

$$G_2 = \boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{g}_1 + \sqrt{\kappa}W, \quad (32)$$

where $W \sim \mathcal{N}(0, 1)$. For the Cauchy case, from (27), we note that the variable $X_2|\vec{X}_1 = \mathbf{x}_1$ can be expressed as

$$X_2 = \underbrace{\boldsymbol{\Sigma}_{12}^{\mathsf{T}}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1}_{\hat{\mu}} + \sqrt{\frac{\kappa(1 + \|\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|^2)}{d}}W, \quad (33)$$

where W is a standard Student's t random variable with $\nu = d$, i.e., $W \sim \mathcal{T}(d, 1, 0)$. The similarity between (32) and (33) is apparent. On the other hand, the difference is in the distribution of W and its coefficient, the latter of which is dependent on $\mathbf{x}_1$ in (33). The above expressions offer an effective method to generate random variates for $G_2|\vec{G}_1 = \mathbf{g}_1$ and $X_2|\vec{X}_1 = \mathbf{x}_1$ (for $\alpha = 1$), respectively. However, for general α, one *cannot* express the PDF of $X_2|\vec{X}_1 = \mathbf{x}_1$ in closed-form. Therefore, one must revert to alternate methods to generate the corresponding random variates.

4.1. Random Variates via Rejection Sampling

Rejection sampling offers us a generic framework for random variate generation from an arbitrary PDF [19]. In our case, we want to generate outcomes for the conditional random variable $X_2|\vec{X}_1 = \mathbf{x}_1$, the PDF of which is $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$. To employ the rejection sampling algorithm, a PDF $\rho(x_2)$ needs to be specified such that $c\rho(x_2)$ *majorizes* $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ for some $c \geq 1$, i.e., $c\rho(x_2) \geq f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) \,\forall\, x_2 \in \mathbb{R}$. The idea is to find a $\rho(\cdot)$ from which it is *sufficiently easy* to generate random variates. At the same time, $\rho(x_2)$ is desired to approximate $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ as close as possible for all x_2, i.e., $c \approx 1$. A summary of the algorithm is as follows: Generate h from $\rho(\cdot)$, u from $\mathcal{U}(0, 1)$ and check if $cu \leq f(h|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)/\rho(h)$. If the condition is satisfied, accept $X_2 = h$. Else reject the sample and repeat the procedure [19, pg. 51].

We note that both $\rho(\cdot)$ and c need to be determined before the algorithm can be implemented. We address each of these issues below:

4.1.1. Choice of $\rho(\cdot)$

From the discussion in Section 3, it is not hard to see why selecting a Student's t-distribution for $\rho(\cdot)$ would be appropriate. Besides being asymptotically equivalent as highlighted in Proposition 1, we note that the tails of the PDFs in (21) and (26) decay as a power law. One may equate these orders to evaluate a suitable value for ν. A straightforward derivation in Appendix B gives us

$$\lim_{x_2 \to \infty} x_2^{\alpha+d} f(x_2 + \hat{\mu}|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d) = c_1, \quad (34)$$

while from (26) we have

$$\lim_{x_2 \to \infty} x_2^{\nu+1} f_S(x_2 + \mu; \nu, \delta, \mu)$$
$$= \lim_{x_2 \to \infty} x_2^{\nu+1} f_S(x_2; \nu, \delta, 0) = c_2, \quad (35)$$

where c_1 and c_2 are constants independent of x_2. On comparing (34) and (35), we note that the tail order of $f(x_2|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d)$ is tracked by $\nu = \hat{\nu}$ in (35). Likewise, as $f_S(x_2; \hat{\nu}, \delta, \mu)$ is unimodal and symmetric about μ, equating the latter to $\hat{\mu}$ is appropriate. Setting the scale parameter is a matter of perspective. One possible method involves noting that

$$f_S(x_2; \hat{\nu}, \delta, \hat{\mu}) = \frac{1}{\delta} f_S\left(\frac{x_2 - \hat{\mu}}{\delta}; \hat{\nu}, 1, 0\right)$$

and enforcing the constraint

$$f_S(\hat{\mu}; \hat{\nu}, \hat{\delta}, \hat{\mu}) = f(\hat{\mu}|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d) \quad (36)$$
$$\Rightarrow \hat{\delta} = \frac{f_S(0; \hat{\nu}, 1, 0)}{f(\hat{\mu}|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d)}. \quad (37)$$

The equality in (36) ensures that $f(x_2|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d)$ and $f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu})$ are equivalent at the mode $\hat{\mu}$ of either distribution. On substituting (21) and (26) in (37), we get

$$\hat{\delta} = \sqrt{\frac{\kappa}{\hat{\nu}}} \frac{\Gamma(\frac{\hat{\nu}+1}{2})\Gamma(\frac{d-1}{2})}{\Gamma(\frac{\hat{\nu}}{2})\Gamma(\frac{d}{2})} \frac{v(\|\mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d-1)}{v(\|\mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d)}. \quad (38)$$

Due to (36) and Proposition 1, we also note that

$$f(x_2|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d) \sim f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu}) \quad (39)$$

$\forall \, x_2 \in \mathbb{R}$ as $\|\mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}_1\| \to \infty$, thus justifying the choice of $\rho(x_2) = f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu})$ in an asymptotic sense.

Finally, we note that generating random variates from a Student's t-distribution is computationally simple to do [20]. A single sample is derived from a Gaussian and a chi-squared random variable. The degrees of freedom for the latter is equivalent to that of the Student's t-distribution.

4.1.2. Evaluating c

As the probability of acceptance in rejection sampling is $1/c$, it is desired that c be as close to 1 as possible. This can be done numerically. From
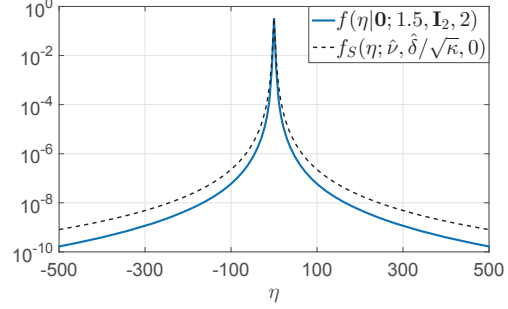

Figure 3: A comparison of $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_2, 2)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$.
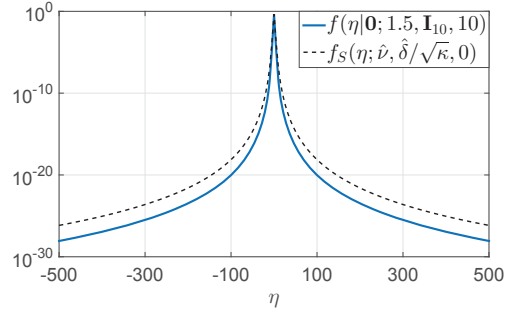

Figure 4: A comparison of $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_{10}, 10)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$.

[19], we know that c needs to satisfy the inequality

$$c f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu}) \geq f(x_2|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d) \quad (40)$$

$\forall \, x_2 \in \mathbb{R}$. On substituting (22) in (40) and invoking the scale and shift properties of $f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu})$ we have

$$c f_S\left(\frac{x_2 - \hat{\mu}}{\sqrt{\kappa}}; \hat{\nu}, \frac{\hat{\delta}}{\sqrt{\kappa}}, 0\right) \geq f\left(\frac{x_2 - \hat{\mu}}{\sqrt{\kappa}}\Big|\hat{\mathbf{x}}_1; \alpha, \mathbf{I}_d, d\right). \quad (41)$$

We note that if c satisfies the condition

$$c \geq \lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d) \quad (42)$$

where

$$\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d) = \frac{f(\eta|\hat{\mathbf{x}}_1; \alpha, \mathbf{I}_d, d)}{f_S\left(\eta; \hat{\nu}, \frac{\hat{\delta}}{\sqrt{\kappa}}, 0\right)} \quad (43)$$

$\forall \, \eta \in \mathbb{R}$, then (40) holds as well. To avoid evaluating c for every $\|\hat{\mathbf{x}}_1\|$, we would want to maximize
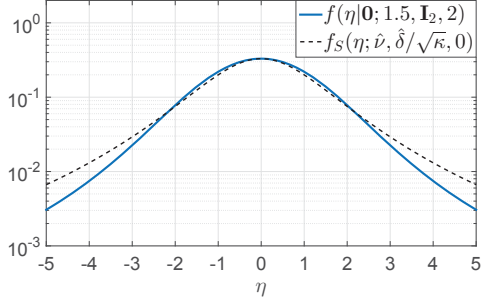
Figure 5: A comparison of $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_2, 2)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for η close to zero.
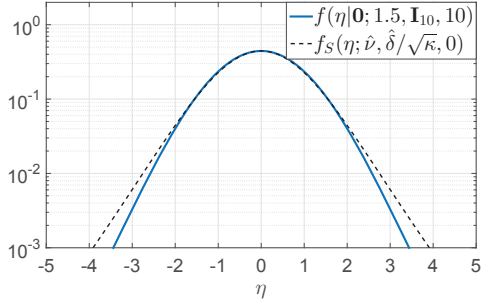

Figure 6: A comparison of $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_{10}, 10)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for η close to zero.

(43) over both η and $\|\hat{\mathbf{x}}_1\|$ for a given α and d. The optimal c in this sense is

$$\hat{c}(\alpha, d) = \max_{\eta, \|\hat{\mathbf{x}}_1\|} \lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d). \qquad (44)$$

As both $f(\eta|\hat{\mathbf{x}}_1; \alpha, \mathbf{I}_d, d)$ and $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ are symmetric about η and equivalent at $\eta = 0$, we may evaluate (44) for $\eta \in \mathbb{R}^+$. Note that for any $\boldsymbol{\Sigma}$, we can simplify the initial problem in (40) to (44), the latter of which depends on the conditional PDF of a *standard isotropic* random αSG vector. For comparison we present $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_d, d)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for $d = 2$ and $d = 10$ in Figs. 3 and 4, respectively. Clearly, the conditional PDFs in either figure are tracked well by their Student's t counterparts. Though Figs. 3 and 4 offer us sufficient motivation to employ Student's t-PDFs as good majorizing functions, we need to make some sense of where the maxima of $\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d)$ lies in the η-$\|\hat{\mathbf{x}}_1\|$ plane. By doing so, we are able to quickly evaluate $\hat{c}(\alpha, d)$ by just analyzing small regions in η-$\|\hat{\mathbf{x}}_1\|$.
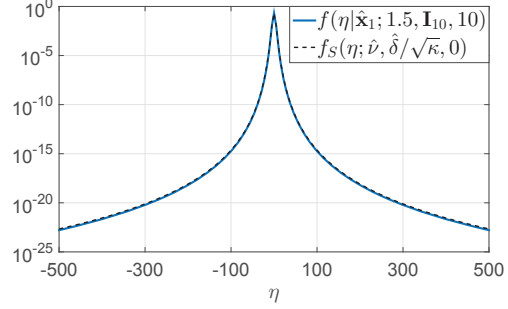

Figure 7: A comparison of $f(\eta|\hat{\mathbf{x}}; 1.5, \mathbf{I}_{10}, 10)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for $\|\hat{\mathbf{x}}_1\|^2 = 40$.

To highlight the trends near $\eta = 0$, we plot $f(\eta|\mathbf{0}; 1.5, \mathbf{I}_d, d)$ and the corresponding $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for $d = 2$ and $d = 10$ in Figs. 5 and 6, respectively. In both cases we note that the Student's t-PDFs have relatively sharper peaks than the corresponding conditional PDFs. One would therefore expect (44) to exist at *small values* of η. For the instances in Figs. 5 and 6, we can clearly see that $\hat{c}(\alpha, d)$ exists in the interval $0 < \eta < 2$. To see the impact of $\|\hat{\mathbf{x}}_1\|$ on (44), from Proposition 1, we note that (42) is satisfied for $c = 1$ as $\|\hat{\mathbf{x}}_1\| \to \infty$. To highlight this, we present $f(\eta|\hat{\mathbf{x}}_1; 1.5, \mathbf{I}_d, d)$ and $f_S(\eta; \hat{\nu}, \hat{\delta}/\sqrt{\kappa}, 0)$ for $d = 2$ and $\|\hat{\mathbf{x}}_1\|^2 = 40$ in Fig. 7. Compared with Fig. 4, the PDFs are much closer together in Fig. 7 and would thus satisfy (42) for c close to 1. Consequently, $\hat{c}(\alpha, d)$ should exist at *small values* of $\|\hat{\mathbf{x}}_1\|$. In fact, from $f(\eta|\hat{\mathbf{x}}_1; 1.5, \mathbf{I}_d, d)$ and (21), $\|\hat{\mathbf{x}}_1\|$ should be such that $v(\|\hat{\mathbf{x}}_1\|; \alpha, d - 1)$ and $v((\|\hat{\mathbf{x}}_1\|^2 + \eta^2)^{1/2}; \alpha, d)$ *do not* operate in their respective linear descending regions, as this is where one expects the maximum divergence between the PDFs in (43).

To further strengthen the above arguments, we consider all possible (α, d) pairs for $1.1 \leq \alpha \leq 1.99$ (with a precision of two decimal places) and $d \in \{1, 2, \ldots, 10\}$. These parameter ranges exemplify the statistics found in typical impulsive data sets [6, 7, 14]. For each (α, d), we determine $(\eta, \|\hat{\mathbf{x}}_1\|)$ that maximizes $\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d)$ over all possible pairings of $\eta \in \{0, 0.1, \ldots, 2\}$ and $\|\hat{\mathbf{x}}_1\| \in \{0, 0.2, \ldots, 2\}$. Thereafter, the bin count corresponding to the determined coordinate is increased by one. In Fig. 8, we present the accumulated bin count which is shown as a color map in the η-$\|\hat{\mathbf{x}}_1\|$ plane. We note that for the consid-
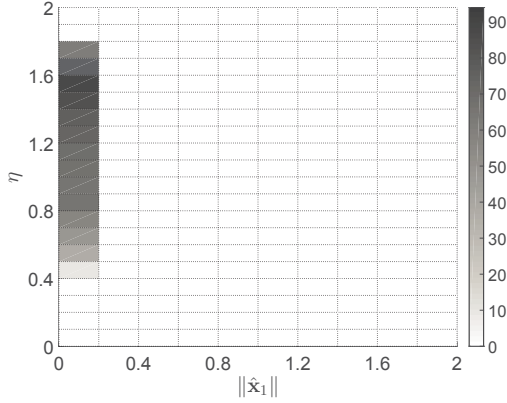
9

Figure 8: Bin counts of `max` $\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d)$ in the η-$\|\hat{\mathbf{x}}_1\|$ plane for $1.1 \leq \alpha \leq 1.99$ (with a precision of two decimal places) and $d \in \{1, 2, \ldots, 10\}$.

Table 1: $\hat{c}(\alpha, d)$ rounded to the nearest hundredth for select values of α and d.

		d				
		2	4	6	8	10
	1.1	1.03	1.02	1.02	1.02	1.01
	1.2	1.06	1.04	1.03	1.03	1.02
	1.3	1.08	1.05	1.04	1.03	1.03
	1.4	1.11	1.07	1.05	1.04	1.03
α	1.5	1.13	1.08	1.06	1.05	1.04
	1.6	1.14	1.09	1.06	1.05	1.04
	1.7	1.16	1.10	1.07	1.06	1.05
	1.8	1.17	1.11	1.08	1.06	1.05
	1.9	1.19	1.11	1.08	1.06	1.05

ered range and precision of α and d, the maxima of $\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d)$ is found within $0.4 \leq \eta \leq 1.8$ and $0 \leq \|\hat{\mathbf{x}}_1\| \leq 0.2$. Fig. 8 offers us some sense of the distribution of $\hat{c}(\alpha, d)$ in the η-$\|\hat{\mathbf{x}}_1\|$ plane. If $\hat{c}(\alpha, d)$ is to be determined, then for most practical purposes ($1.1 \leq \alpha < 2$ and $d \leq 10$), one may just maximize (44) over the intervals $0 \leq \|\hat{\mathbf{x}}_1\| \leq 0.2$ and $0.4 \leq \eta \leq 1.8$ instead of the entire η-$\|\hat{\mathbf{x}}_1\|$ plane.

In Table-1 we highlight numerically evaluated values of $\hat{c}(\alpha, d)$, rounded to the largest hundredth for *select instances* of $\alpha \geq 1$ and d. From the insights gained from Fig. 8, the values are determined by calculating $\lambda(\eta, \|\hat{\mathbf{x}}_1\|, \alpha, d)$ for random samples of $0.4 \leq \eta \leq 1.8$ and $0 \leq \|\hat{\mathbf{x}}_1\| \leq 0.2$, selecting the maximum and rounding it off to the nearest largest hundredth. We see that $\hat{c}(\alpha, d)$ decreases with d and increases with α. Overall, a number of results were compiled for $1.1 \leq \alpha \leq 1.99$ and $d \leq 10$. We found $c = 1.2$ to be sufficient to satisfy (42) within this range.

Remark 3. From the discussion in Sections 4.1.1 and 4.1.2, we note that *any* appropriately shifted (by $\hat{\mu}$) and scaled *symmetric* PDF with tails heavier than $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ can be used as $\rho(\cdot)$. Though such a function may offer a low c for some (α, d) pair, it may not perform as well for others. On the other hand, $f_S(x_2; \hat{\nu}, \hat{\delta}, \hat{\mu})$ tracks $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ $\forall x_2 \in \mathbb{R}$ very well for practical ranges of α and d. This is evident from Table 1. Further still, it tracks the tail order of $f(x_2|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)$ and is asymptotically equivalent to the latter (via Proposition 1). Finally, generating random variates from the Stu-

dent's t-distribution is a straight forward process [20]. These properties make the Student's t-PDF desirable to be used in the rejection sampling algorithm.

4.2. Application & Computational Aspects

In certain applications, one may need to generate random variates from the conditional distribution in (21). Simulations involving models that employ sliding windows and constrain windowed samples to follow a multivariate αSG distribution fall in this category. An example is the generation of realizations from the αSGN(m) process [7] which was introduced to model the elliptic dependence structures found within delay scatter plots of impulsive noise data in warm shallow waters. The noise model uses a sliding window of length $m+1$ and constrains the samples within to be αSG. By further constraining the underlying covariance matrix and α to be time invariant, stationarity is guaranteed.

To generate an αSGN(m) realization, the noise samples are returned sequentially, each from a PDF conditional on the last m samples. Let $\vec{X}_{t,m} = [X_{t-m}, X_{t-m+1} \ldots, X_t]^{\mathsf{T}}$ be the random vector corresponding to $m + 1$ such samples at time t, then one can extend (3) to

$$\vec{X}_{t,m} = \sqrt{A_t}\vec{G}_{t,m}, \qquad (45)$$

where $\vec{G}_{t,m} = [G_{t-m}, G_{t-m+1} \ldots, G_t]^{\mathsf{T}}$ is the underlying Gaussian vector at time t with covariance matrix $\boldsymbol{\Sigma}$. Moreover, due to the sliding window framework and stationarity of the process, $\boldsymbol{\Sigma}$ is now a *symmetric Toeplitz matrix* [7]. From (32),

$G_t | \vec{G}_{t-1,m-1} = \mathbf{g}_{t-1,m-1}$ can be expressed as

$$G_t = \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{g}_{t-1,m-1} + \sqrt{\kappa} W_t, \qquad (46)$$

where $W_t \sim \mathcal{N}(0,1)$ are IID noise samples for all $t \in \mathbb{Z}$. Similarly, from (33), we can express $X_t | \vec{X}_{t-1,m-1} = \mathbf{x}_{t-1,m-1}$ as

$$X_t = \boldsymbol{\Sigma}_{12}^{\mathsf{T}} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{x}_{t-1,m-1}$$
$$+ \sqrt{\frac{\kappa(1 + \|\boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{x}_{t-1,m-1}\|^2)}{d}} W_t, \quad (47)$$

for the Cauchy case, where $W_t \sim \mathcal{T}(d,1,0)$ are IID samples for all $t \in \mathbb{Z}$. One notes that (46) is an mth-order autoregressive (AR(m)) process [17], however (47) is not as the coefficient of W_t is time-varying due to its dependence on $\mathbf{x}_{t-1,m-1}$. Therefore, (47) is a non-linear generalized autoregressive conditional heteroskedasticity (GARCH) process [1]. For general α, one can use rejection sampling to generate the noise realization. This is summed up in Algorithm 1. To initialize the process, one generates $\mathbf{x}_{m,m-1} = \mathbf{x}_0$ from the d-dimensional αSG distribution with underlying covariance matrix $\boldsymbol{\Sigma}$. This can be done easily via (3): Generate $A \sim \mathcal{S}(\frac{\alpha}{2}, 1, 2(\cos(\frac{\pi\alpha}{4}))^{2/\alpha}, 0)$ and $\vec{G} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{11})$, then $\mathbf{x}_0 = \sqrt{A}\vec{G}$.

Though optimized due to the choice of $\rho(\cdot)$ and c, Algorithm 1 is still computationally intensive. For each x_t, we require $f(h|\hat{\mathbf{x}}_1; \alpha, \mathbf{I}_d, d)$, which depends on the two integrals $v((\|\hat{\mathbf{x}}_1\|^2 + h^2)^{1/2}; \alpha, d)$ and $v(\|\hat{\mathbf{x}}_1\|; \alpha, d-1)$ as highlighted by (15). Likewise, from (38), $\hat{\delta}$ requires the further evaluation of $v(\|\hat{\mathbf{x}}_1\|; \alpha, d)$. Consequently, in one iteration of the loop in Algorithm 1, there are *three* integrals that need to be numerically evaluated. From (7) and (5), we note that these functions depend upon the heavy-tailed $f_A(t; \alpha)$. Though numerical methods do exist, the resulting computations are very time consuming [14, 13], especially when N is large. One can avoid this by doing a one-time tabulation of $v(r; \alpha, d)$ and $v(r; \alpha, d-1)$ over r for the values of α and d that are of interest.

The log-log plots in Figs. 1 & 2 clearly show two linear regions, with the transition being roughly between $r = 1$ and $r = 10$. These trends are seen for other values of α and d. Sampling at high resolution for $r \in [10^{-3}, 10^3]$ (in the log-log domain) offers us enough information about the transitional range. One can therefore use linear interpolation

---

Algorithm 1: Generate N samples of αSGN(m).

Input: $\alpha, m, N, \boldsymbol{\Sigma}$
Initialize: Set $\mathbf{x}_{m,m-1} = \mathbf{x}_0$, $d = m+1$ and evaluate $\hat{\nu}$ and c

1 while $m + 1 \leq t \leq N$ **do**
2 $\mathbf{x}_1 \leftarrow \mathbf{x}_{t-1,m-1}$
3   Evaluate $\hat{\delta}$ and $\hat{\mu}$
4   Generate h from $\mathcal{T}(\hat{\nu}, \hat{\delta}, \hat{\mu})$
5   Generate u from $\mathcal{U}(0, 1)$
6   **if** $cu \leq \frac{f(h|\mathbf{x}_1; \alpha, \boldsymbol{\Sigma}, d)}{f_S(h; \hat{\nu}, \hat{\delta}, \hat{\mu})}$ **then**
7 $x_t \leftarrow h$
8 $t \leftarrow t + 1$
9   **end**
10 end

---

between immediately lower and higher sampling indices to evaluate intermediate values of r within this interval. For $r < 10^{-3}$ and $r > 10^3$, the limiting closed-form expressions in (12) and (11) may be used, respectively.

5. Conclusions

In this paper, we derived and discussed various properties of a univariate conditional distribution stemming from a multivariate αSG PDF. The former was shown to be unimodal, heavy-tailed and converged to a Student's t-distribution in an asymptotic sense. Using these insights, we propose a method based on rejection sampling that allowed generating random variates from the conditional distribution efficiently. As the PDF of the latter is not in closed-form and is evaluated by integrating over a skewed heavy-tailed PDF, the algorithm was still computationally taxing. For a given α and d, we show how this could be avoided by performing a one-time tabulation of three functions.

Appendix A. Derivation of (16)

We can represent the lower-triangular matrix $\boldsymbol{\Sigma}^{1/2}$ in the block form

$$\boldsymbol{\Sigma}^{1/2} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{12}^{\mathsf{T}} & \sqrt{\kappa} \end{bmatrix}, \qquad (A.1)$$

where $\mathbf{L}_{11} \in \mathbb{R}^{(d-1)\times(d-1)}$ is a lower-triangular matrix, $\mathbf{L}_{12} \in \mathbb{R}^{(d-1)\times 1}$ and κ is a scalar. Therefore,

$$\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}^{1/2})^{\mathsf{T}} = \begin{bmatrix} \mathbf{L}_{11}\mathbf{L}_{11}^{\mathsf{T}} & \mathbf{L}_{11}\mathbf{L}_{12} \\ \mathbf{L}_{12}^{\mathsf{T}}\mathbf{L}_{11}^{\mathsf{T}} & \mathbf{L}_{12}^{\mathsf{T}}\mathbf{L}_{12} + \kappa \end{bmatrix}. \tag{A.2}$$

Comparing this with (13) leads to the following expressions:

$$\mathbf{\Sigma}_{11} = \mathbf{L}_{11}\mathbf{L}_{11}^{\mathsf{T}} \tag{A.3}$$

$$\mathbf{\Sigma}_{12} = \mathbf{L}_{11}\mathbf{L}_{12} \tag{A.4}$$

$$\sigma_{22}^2 = \mathbf{L}_{12}^{\mathsf{T}}\mathbf{L}_{12} + \kappa. \tag{A.5}$$

Clearly, $\mathbf{L}_{11}$ arises from the Cholesky decomposition of $\mathbf{\Sigma}_{11}$, i.e., $\mathbf{L}_{11} = \mathbf{\Sigma}_{11}^{1/2}$. As $\mathbf{\Sigma}_{11}$ is of full rank, we have from (A.4), $\mathbf{L}_{12} = \mathbf{L}_{11}^{-1}\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}$. Finally from (A.5), $\kappa = \sigma_{22}^2 - \mathbf{L}_{12}^{\mathsf{T}}\mathbf{L}_{12} = \sigma_{22}^2 - \mathbf{\Sigma}_{12}^{\mathsf{T}}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$, which is the Schur's complement of $\mathbf{\Sigma}_{11}$ [18, pg. 47]. Substituting these forms back in (A.1) gives us (16).

Appendix B. Derivation of (34)

From (19), we note that

$$\lim_{x_2 \to \infty} \|\mathbf{\Sigma}^{-1/2}\mathbf{x}\| / (x_2 - \hat{\mu}) = 1/\sqrt{\kappa}. \tag{B.1}$$

On substituting (B.1) in (15), we have

$$\lim_{x_2 \to \infty} \frac{f(x_2 + \hat{\mu}|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d)}{v\left(x_2/\sqrt{\kappa}; \alpha, d\right)}$$
$$= \frac{\Gamma(\frac{d}{2})}{\sqrt{\kappa\pi}\,v(\|\mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d-1)\Gamma(\frac{d-1}{2})}. \tag{B.2}$$

As $x_2 \to \infty$, one can invoke the asymptotic equality in (11) for $v(x_2/\sqrt{\kappa}; \alpha, d)$ in (B.2) to get

$$\lim_{x_2 \to \infty} x_2^{\alpha+d} f(x_2 + \hat{\mu}|\mathbf{x}_1; \alpha, \mathbf{\Sigma}, d) = c_1, \tag{B.3}$$

where

$$c_1 = \frac{\alpha k_1(\alpha, d)\kappa^{(\alpha+d-1)/2}}{\sqrt{\pi}\,v(\|\mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}_1\|; \alpha, d-1)}\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})}.$$

References

[1] Bollerslev, T., 2008. Glossary to ARCH (GARCH). CRE-ATES Research Paper 49.

[2] Chitre, M. A., Potter, J. R., Ong, S.-H., Apr 2006. Optimal and Near-Optimal Signal Detection in Snapping Shrimp Dominated Ambient Noise. IEEE J. of Ocean. Eng. 31 (2), 497–503.

[3] Fertonani, D., Colavolpe, G., Jul 2009. On reliable communications over channels impaired by bursty impulse noise. IEEE Trans. on Commun. 57 (7), 2024–2030.

[4] Legg, M. W., 2009. Non-Gaussian and non-homogeneous Poisson models of snapping shrimp noise. Curtin Univ. of Technology.

[5] Lin, J., Nassar, M., Evans, B. L., 2013. Impulsive noise mitigation in powerline communications using sparse bayesian learning. IEEE J. on Sel. Areas in Commun. 31 (7), 1172–1183.

[6] Mahmood, A., Jun 2014. Digital Communications in Additive White Symmetric Alpha-Stable Noise. Ph.D. thesis, Natl. Univ. of Singapore.

[7] Mahmood, A., Chitre, M., May 2015. Modeling colored impulsive noise by markov chains and alpha-stable processes. In: OCEANS 2015 - Genova. pp. 1–7.

[8] MATLAB, 2014. version 8.4.0 (R2014a). The MathWorks Inc., Natick, Massachusetts.

[9] Mohammadpour, A., Soltani, A., 2004. On simulating exchangeable sub-gaussian random vectors. Stat. & Probab. Lett. 69 (1), 29 – 36.

[10] Ndo, G., Labeau, F., Kassouf, M., Oct 2013. A Markov-Middleton Model for Bursty Impulsive Noise: Modeling and Receiver Design. IEEE Trans. on Power Deliv. 28 (4), 2317–2325.

[11] Nikias, C. L., Shao, M., 1996. Signal processing with Alpha-Stable Distributions and Applications. Chapman-Hall, New York.

[12] Nolan, J., 1999. An algorithm for evaluating stable densities in zolotarev's (m) parameterization. Mathematical and Computer Modelling 29 (1012), 229 – 233.

[13] Nolan, J. P., 1997. Numerical calculation of stable densities and distribution functions. Commun. in Stat. Stoch. Models 13 (4), 759–774.

[14] Nolan, J. P., 2013. Multivariate elliptically contoured stable distributions: theory and estimation. Comput. Stat. 28 (5), 2067–2089.

[15] Nolan, J. P., 2015. Stable Distributions - Models for Heavy Tailed Data. Birkhauser, Boston.

[16] Nolan, J. P., 2016.
URL http://academic2.american.edu/~jpnolan/stable/stable.html

[17] Papoulis, A., Pillai, U. S., dec 2001. Probability, Random Variables and Stochastic Processes. McGraw-Hill, Boston.

[18] Petersen, K. B., Pedersen, M. S., nov 2012. The Matrix CookBook.
URL http://www2.imm.dtu.dk/pubdb/p.php?3274

[19] Robert, C., Casella, G., 2013. Monte Carlo Statistical Methods. Springer Texts in Stat. Springer New York.

[20] Roth, M., apr 2013. On the Multivariate t Distribution. Tech. rep., Linkoping university, Division of automatic control.

[21] Samorodnitsky, G., Taqqu, M. S., 1994. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall.

[22] Strang, G., 2006. Linear Algebra and Its Applications. Thomson Brooks/Cole Cengage learning.

[23] Zimmermann, M., Dostert, K., Feb 2002. Analysis and modeling of impulsive noise in broad-band powerline communications. IEEE Trans. on Electromagn. Compat. 44 (1), 249–258.