Spectra of Variance-Stabilized Correlated Noise: Modeling and Efficient Computation

Andrea Corsini and Alessandro Foi
Signal Processing Research Centre
Tampere University, Finland
firstname.lastname@tuni.fi

***Abstract*—Variance-stabilizing transformations (VSTs) are established statistical tools for tackling heteroskedasticity. They are commonly utilized as a preprocessing module within signal-processing pipelines, enabling one to use off-the-shelf methods designed for data corrupted by noise of constant variance on data that is otherwise corrupted by noise with signal-dependent variance. Being univariate mappings, VSTs are agnostic to possible correlation in the data. However, contrary to common assumptions, applying VSTs not only stabilizes the noise variance, but can also significantly perturb the noise correlation hence distorting the noise spectrum. As the accurate knowledge of the noise spectrum is essential to many applications such as filtering and coding, any unpredictable distortion of the noise spectrum is detrimental. In this paper, we propose a method to compute the distorted noise spectrum of the variance-stabilized data. The proposed method is efficient and accurate, extending the scope of reliable application of VSTs to data corrupted by correlated noise.**

***Index Terms*—noise modeling, variance stabilization, correlated noise, noise power spectrum.**

I. Introduction

Real-world data are typically heteroskedastic. Images corrupted by signal-dependent noise are prominent examples of heteroskedastic data, where the noise variance depends on the unknown clean signal. However, most signal processing tools are designed for homoskedastic noise, i.e. having constant variance. Variance-stabilizing transformations (VSTs) [1]–[4] come in handy for preprocessing heteroskedastic data and to make these data amenable to methods meant for homoskedastic data. For instance, Poisson and Poisson-Gaussian data can be stabilized asymptotically for large mean values by the Anscombe VSTs [1], [5] whereas data corrupted by multiplicative noise are stabilized by logarithmic transformations [6], [7]. VSTs find application in a wide range of sophisticated estimation and restoration problems [8], [9] and are beneficial even when used in conjunction with deep neural networks [10], [11].

Noise corrupting real-world data is often also correlated. Correlation can be characterized by the power spectral density (PSD) of the noise, an essential parameter for effective processing [12]. While one may assume that the VST leaves the noise correlation unchanged [13], here we show that VSTs can significantly distort the correlation, and we provide a method to accurately compute the noise PSD after a given VST.

Given a VST and the variance function of Gaussian[1] noise, we locally approximate the VST with a polynomial of arbitrary degree (Section IV-A). Then, exploiting the Isserlis' theorem [14] (Section IV-B), we are able to compute the spectrum of the variance-stabilized signal given the spectrum of the signal before stabilization (Section IV-C). We show that using the polynomial approximation as a surrogate for the VST for the spectrum calculation, we can efficiently and accurately approximate the actual spectrum of the variance-stabilized data obtained from Monte Carlo simulations (Section VI). To the best of our knowledge, this is the very first method that resolves the issue of the spectra of variance-stabilized correlated noise.

II. Background

A. Spatially correlated signal-dependent noise

We consider a degraded signal $z : \Omega \to \mathbb{R}$ consisting of an unknown deterministic clean image $y : \Omega \to Y \subseteq \mathbb{R}$ corrupted by Gaussian signal-dependent noise, where $\Omega \subset \mathbb{Z}^d$ is the signal domain, e.g. $d = 2$ for images. We model the signal as

$$z(x) = y(x) + \sigma(y(x))\eta(x), \quad \eta(x) \sim \mathcal{N}(0,1), \tag{1}$$

where $x \in \Omega$ is the sample coordinate[2]. The conditional mean and variance are $\mathrm{E}\{z(x)|y(x)\} = y(x)$ and $\mathrm{var}\{z(x)|y(x)\} = \sigma^2(y(x))$, respectively. Here, $\sigma : Y \to \mathbb{R}^+$ is called the standard deviation function and expresses how the standard deviation of the noise depends on the mean of the signal. Examples of σ are the square root of affine functions for Poisson-Gaussian data [15], linear functions for Rayleigh data [7], and the square root of quadratic functions for photon-detectors with photo-response nonuniformity [4].

This work considers η in (1) to be stationary spatially correlated having PSD $\Psi = \mathrm{var}\{\mathcal{F}[\eta]\}$, where $\mathcal{F}$ denotes the Fourier transform. This is known as the correlated signal-dependent noise model with *noise-scaling post correlation* [16]. The noise PSD is linked to the autocovariance function by

$$\rho_\Psi(\xi) = \mathrm{E}\{\eta(x)\eta(x+\xi)\} = \mathcal{F}^{-1}[\Psi](\xi)|\Omega|^{-1}, \tag{2}$$

[1]This is a reasonable practical assumption given that in acquisition system pipelines there are several noise contributions, and their combination can be considered Gaussian by virtue of the central limit theorem.

[2]We omit the coordinate x from $z(x)$ and $y(x)$ to streamline the notation where possible.

where $\xi \in \Omega$ is a lag in circular sense. As $\eta(\cdot) \sim \mathcal{N}(0,1)$, we always have that $\|\Psi\|_1 = |\Omega|^2$, where $|\Omega|$ is the cardinality of the signal domain, and $\rho_\Psi(0) = 1$. In the particular case of white noise, we also have $\Psi = |\Omega|$ and $\rho_\Psi(\xi) = 0$ for $\xi \neq 0$.

The PSD of z can be approximated as $\mathrm{var}\{\mathcal{F}[z]\} \approx \Psi \|\sigma^2(y)\|_1 |\Omega|^{-1}$ [16]. This approximation is exact when the noise on z is stationary, i.e. when y and, in turn, $\sigma(y)$ are constant. In regions with y approximately constant, z is approximately stationary and the PSD of z over such regions can be given by scaling Ψ, i.e. $\Psi\sigma^2(y)$.

B. Variance stabilization

A variance stabilizing transformation for z (1) is a function $f : \mathbb{R} \to \mathbb{R}$ such that $\mathrm{std}\{f(z)|y\} \approx c$, where $c > 0$ is a constant, thus making the conditional standard deviation of $f(z)$ independent of y. In the following, we assume $c = 1$ without loss of generality.

When using a VST, it is commonly assumed that the stabilization is exact. Another common assumption is that f does not distort the spatial correlation of the noise, which implies that the PSD of $f(z)$ can be given as a rescaling of the PSD of z [13] for which $\mathrm{var}\{f(z)\} = 1$, i.e. $\mathrm{var}\{\mathcal{F}[f(z)]\} = \Psi$.

III. Problem

The motivation of this paper is in the known failures of the aforementioned assumptions. Firstly, VSTs with imperfect stabilization will not give unitary variance for each value of y, for example the Anscombe transformation [1] stabilizes poorly for small values of y. Secondly, being VSTs nonlinear functions, the PSD of the stabilized noise on $f(z)$ can be different from $\gamma\Psi$ for any constant $\gamma \in \mathbb{R}^+$, due to the spatial correlation on η being distorted.

The contribution of this paper lies in the practical computation of the conditional spectrum of variance-stabilized signals $\mathrm{var}\{\mathcal{F}[f(z)]|y\}$ for any y provided known $\sigma(y)$ and Ψ.

IV. Method

As commented in Section II-A, we consider regions where z is locally stationary, and so is $f(z)$ in these regions. We propose to compute a polynomial approximation of f locally for a neighborhood of noisy data distributed about any given $y \in Y$. The polynomial is used as a surrogate for f to determine the distortion of the spectrum which, as per (2), can be represented in terms of the autocovariance function of z. In particular, we show that the autocovariance function of $f(z)$ can be computed as a special polynomial in the autocovariance of z, whose coefficients are in turn conveniently expressed from the coefficients of the polynomial approximating f.

A. Fitting

Let $p_{\boldsymbol{\alpha}}(t) = \sum_{j=0}^{N} \alpha_j t^j$ be the polynomial of degree N with coefficients $\boldsymbol{\alpha} = (\alpha_j)_{j=0}^{N} \in \mathbb{R}^{N+1}$. Given $y \in Y$, the weighted least-squares N^{th}-order local polynomial approximation of f at y has coefficients $\boldsymbol{\alpha}^\star(y)$ such that

$$\boldsymbol{\alpha}^\star(y) = \arg\min_{\boldsymbol{\alpha}} \int_{\mathbb{R}} \left(f(t) - p_{\boldsymbol{\alpha}}(t)\right)^2 w_y(t)\,\mathrm{d}t \qquad (3)$$

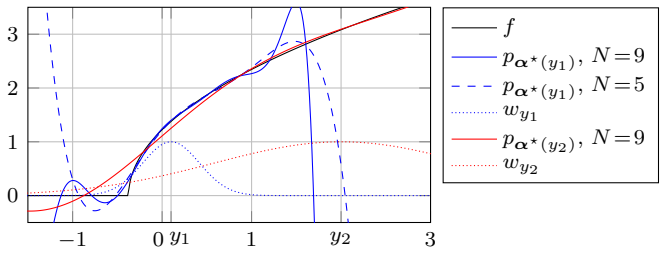

Fig. 1. Local polynomial approximation of $f(z) = 2\sqrt{\max\left(0, z + \frac{7}{8}\right)}$ at $y_1 = 0.1$ and $y_2 = 2$ with $\sigma(y) = \sqrt{y}$. The dotted lines show the weights w_{y_i}, $i = 1, 2$, and the solid lines show the corresponding fitted polynomial $p_{\boldsymbol{\alpha}^\star(y_i)}$ of degree $N = 9$. A dashed line shows also a fitting at y_1 for $N = 5$. Notice the poorer fit of the polynomials outside the range $y_i \pm 3\sigma(y_i)$. We stress that $p_{\boldsymbol{\alpha}^\star(y_i)}$ are solutions of (3) and are not Taylor expansions of f, and therefore they may not intersect f at y.

where w_y are weights that define the localization around y. We specifically use weights $w_y(t) = e^{-\frac{1}{2}\left(\frac{t-y}{\sigma(y)}\right)^2}$ that reflect the conditional distribution of $z|y$ (1), so that the polynomial fit is tighter or looser according to the conditional probability of $z - y$ given y.

Fig. 1 illustrates the polynomial fit $p_{\boldsymbol{\alpha}^\star(y)}$ for a given f and $\sigma(y)$ at two distinct values of y. The figure also shows the weights w_y used for the two fittings.

B. Covariance

It is convenient to rewrite $p_{\boldsymbol{\alpha}^\star(y)}(z)$ as a polynomial in η:

$$p_{\boldsymbol{\alpha}^\star(y)}(z) = p_{\boldsymbol{\alpha}^\star(y)}\left(y + \sigma\left(y\right)\eta\right) = p_{\boldsymbol{\beta}^\star(y)}(\eta), \qquad (4)$$

where $\beta_j^\star(y) = \sigma^j(y) \sum_{l=j}^{N} \alpha_l^\star(y) \binom{l}{j} y^{l-j}$. Therefore, given two random variables $z_1 = y + \sigma(y)\eta_1$ and $z_2 = y + \sigma(y)\eta_2$ with $\eta_1, \eta_2 \sim \mathcal{N}(0,1)$ and $\mathrm{cov}\left[\eta_1, \eta_2\right] = \rho$, we can give the covariance between $p_{\boldsymbol{\alpha}^\star}(z_1)$ and $p_{\boldsymbol{\alpha}^\star}(z_2)$ as

$$\begin{aligned}\mathrm{cov}\left[p_{\boldsymbol{\alpha}^\star(y)}(z_1), p_{\boldsymbol{\alpha}^\star(y)}(z_2)\right] &= \mathrm{cov}\left[p_{\boldsymbol{\beta}^\star(y)}(\eta_1), p_{\boldsymbol{\beta}^\star(y)}(\eta_2)\right] \\ &= \sum_{j=1}^{N}\sum_{k=1}^{N} \beta_j^\star(y)\beta_k^\star(y)\mathrm{cov}\left[\eta_1^j, \eta_2^k\right] \\ &= \sum_{j=1}^{N}\sum_{k=1}^{N} \beta_j^\star(y)\beta_k^\star(y)\left(\mathrm{E}\{\eta_1^j\eta_2^k\} - \mathrm{E}\{\eta_1^j\}\,\mathrm{E}\{\eta_2^k\}\right). \qquad (5)\end{aligned}$$

Here the summations can start from $j, k = 1$ even if $\beta_0^\star(y)$ were nonzero. This is because whenever j or k is equal 0, then $\mathrm{cov}\left[\eta_1^j, \eta_2^k\right] = 0$ as at least one of its arguments reduces to a Dirac delta.

To compute (5), we begin from noting that, as $\eta_1 \sim \mathcal{N}(0,1)$, the non-central moment $\mathrm{E}\{\eta_1^j\} = (j-1)!!$ for even j or it is 0 for odd j, and analogous for $\mathrm{E}\{\eta_2^k\}$. For the joint moments $\mathrm{E}\{\eta_1^j\eta_2^k\}$, we can exploit the Isserlis' theorem [14] to have the following convenient recursive formula for $j \geq 2$ and $k \geq 1$:

$$\begin{aligned}\mathrm{E}\{\eta_1^j\eta_2^k\} &= (j-1)\,\mathrm{E}\{\eta_1^2\}\,\mathrm{E}\{\eta_1^{j-2}\eta_2^k\} \\ &\quad + k\,\mathrm{E}\{\eta_1\eta_2\}\,\mathrm{E}\{\eta_1^{j-1}\eta_2^{k-1}\} \\ &= (j-1)\,\mathrm{E}\{\eta_1^{j-2}\eta_2^k\} + k\rho\,\mathrm{E}\{\eta_1^{j-1}\eta_2^{k-1}\}. \qquad (6)\end{aligned}$$

To initialize the recursion and to reduce the amount of actual calculations we note that $\mathrm{E}\{\eta_1^0\eta_2^0\} = 1$, $\mathrm{E}\{\eta_1\eta_2\} = \mathrm{cov}\,[\eta_1, \eta_2] = \rho$, $\mathrm{E}\{\eta_1^j\eta_2^k\} = \mathrm{E}\{\eta_1^k\eta_2^j\}$, and that $\mathrm{E}\{\eta_1^j\eta_2^k\} = 0$ for any odd values of $j + k$.

The recursion (6) exposes the polynomial dependency of (5) on ρ. It is thus appropriate to denote (5) as a function $R : [-1, 1] \times Y \to \mathbb{R}$ of both ρ and y:

$$R(\rho, y) = \mathrm{cov}\left[p_{\boldsymbol{\alpha}^\star(y)}(z_1), p_{\boldsymbol{\alpha}^\star(y)}(z_2)\right]. \tag{7}$$

C. Spectrum

We now consider the autocovariance function (2) $\rho_\Psi : \Omega \to [-1, 1]$ of η assuming the special case of z (1) having constant mean y over Ω. By composing ρ_Ψ with (7), we obtain the autocovariance function of $p_{\boldsymbol{\alpha}^\star(y)}(z)$:

$$R(\rho_\Psi, y)(\xi) = \mathrm{cov}\left[p_{\boldsymbol{\alpha}^\star(y)}(z(x+\xi)), p_{\boldsymbol{\alpha}^\star(y)}(z(x))\right], \tag{8}$$

where $\boldsymbol{\alpha}^\star(y)$ (3) are the polynomial coefficients for any $y \in Y$, as explained in Section IV-A. The PSD of $p_{\boldsymbol{\alpha}^\star(y)}(z)$ can be computed through the Fourier transform of $R(\rho_\Psi, y)$ (8) as

$$\mathcal{F}[R(\rho_\Psi, y)]\,|\Omega| = \mathcal{F}\left[R(\mathcal{F}^{-1}[\Psi]\,|\Omega|^{-1}, y)\right]|\Omega|. \tag{9}$$

We can use (9) as the PSD of $f(z)$ for a generic z (without constant mean) on regions that share a common value of y.

V. Implementation

Let y be such that $\sigma(y) > 0$, as otherwise there is no noise. To compute $\boldsymbol{\alpha}^\star(y)$, we discretize the integral in (3) by approximating it as a summation over a column vector $\boldsymbol{t}_y$ of n evenly spaced points between $y - \Delta\sigma(y)$ and $y + \Delta\sigma(y)$, where Δ is a large enough quantile of the $\mathcal{N}(0, 1)$ distribution. The minimization of the summation has the closed-form solution

$$\boldsymbol{\alpha}^\star(y) = U_y f(\boldsymbol{t}_y), \quad U_y = \left(M_y (M_y^\mathsf{T} W M_y)^{-1}\right)^\mathsf{T} W,$$

where $M_y = (\boldsymbol{t}_y^j)_{j=0}^N$ is the $n \times (N+1)$ frame matrix[3] of monomials, W is a diagonal matrix with the vector $w_y(\boldsymbol{t}_y) = e^{-\frac{1}{2}\left(\frac{t_y - y}{\sigma(y)}\right)^2}$ along the diagonal, and w_y is defined as in Section IV-A.

Notice that W does not depend on y and $\sigma(y)$, as $\frac{\boldsymbol{t}_y - y}{\sigma(y)} = \Delta(2\frac{j}{n-1} - 1)_{j=0}^{n-1}$. Likewise, to avoid computing a different U_y for any different y, we define $\check{\boldsymbol{t}} = \frac{1}{\delta}\frac{\boldsymbol{t}_y - y}{\sigma(y)} = \frac{\Delta}{\delta}(2\frac{j}{n-1} - 1)_{j=0}^{n-1}$ again independent of y and $\sigma(y)$, and set $\check{M} = (\check{\boldsymbol{t}}^j)_{j=0}^N$, and $\check{U} = D(\check{M}(\check{M}^\mathsf{T} W \check{M})^{-1})^\mathsf{T} W$, where D is a diagonal matrix with $(\delta^{-j})_{j=0}^N$ along the diagonal and $\delta > 0$; then for any y, $\boldsymbol{\beta}^\star(y) = \check{U} f(\boldsymbol{t}_y)$. Remarkably, this $\boldsymbol{\beta}^\star(y)$ can be directly plugged into (5) for a straightforward calculation of $R(\rho, y)$ bypassing altogether $\boldsymbol{\alpha}^\star(y)$. Here, setting $\delta = \Delta$ acts as a simple preconditioner to improve the numerical stability for large N.

[3] As per the fundamental theorem of algebra, any $n > N$ ensures the linear independence of the monomials and the invertibility of $M_y^\mathsf{T} W_y M_y$. In practice we have $n \ggg N$; specifically, we use $n = 10\,000$, $N \leq 46$, and $\Delta = 9$.

VI. Experiments

We illustrate the proposed method on two examples of standard deviation functions σ and the corresponding VSTs.

The first example consists of the model for data corrupted by signal-dependent noise with $\sigma(y) = \sqrt{y}$, i.e. the Gaussian approximation of Poisson noise. We stabilize these data by the Anscombe[4] VST $f(z) = 2\sqrt{\max\left(0, z + \frac{7}{8}\right)}$.

The second example consists of data corrupted by noise having affine variance, i.e. $\sigma(y) = \sqrt{ay + b}$, $a = 1/15$, $b = 1/5$, and further subject to clipping [3]. It describes observations captured by a sensor with finite range, which is a typical model for raw images. The clipping of z is expressed as $\tilde{z} = g(z) = \max(0, \min(1, z))$. To stabilize the variance of the noise on $\tilde{z}$, we adopt an optimized VST φ [17]. As $\tilde{z}$ does not follow a Gaussian distribution [15], the Isserlis theorem is not applicable. However, instead of considering $\varphi(\tilde{z})$, where φ stabilizes $\mathrm{var}\{\tilde{z}|y\}$, we can consider the composition of φ and g, and apply our method to $f(z) \coloneqq \varphi(g(z))$, treating this compound f as a VST for z.

A. Examples of covariances for VSTs

We use our method to compute the covariance $R(\rho, y)$ (7) for the two cases of VST f and $\sigma(y)$. We also calculate[5] the ratio between $R(\rho, y)$ and $\rho\,\mathrm{var}\{f(z)|y\}$ to assess the impact of f on ρ. Under the typical assumption that f does not modify the noise covariance and only modifies the variance of z, this ratio should be equal to 1 for all y and ρ.

In the first example, Fig. 2 confirms that the stabilization is inaccurate and begins to get accurate (unitary variance) only as y increases. Moreover, the ratio spans approximately between 1 and 0.9, hence f still distorts the spatial correlation.

In the second example, as visible in Fig. 2, the VST for clipped data does not stabilize accurately near the clipping points $y = 0, 1$, despite f being optimized. The ratio plot shows that the distortion of the correlation caused by f is significant, especially at the clipping points.

Ideally, $R(1, y)$ should match $\mathrm{var}\{f(z)|y\}$, which accordingly we plot over $\rho = 1$ in Fig. 2. We can notice that this is indeed the case in both examples.

B. Monte Carlo for PSDs

For each example, we now verify the accuracy of the proposed method for the calculations of the PSD of $f(z)$ through a Monte Carlo simulation. We simulate many realizations of a stationary z with a given mean y, standard deviation $\sigma(y)$, and a fixed noise PSD Ψ. Provided enough realizations, the sample variance of $\mathcal{F}[f(z)]$ is close to the true $\mathrm{var}\{\mathcal{F}[f(z)]|y\}$.

Fig. 3 reports the results for the Anscombe VST $f(z)$ stabilizing $\sigma(y) = \sqrt{y}$. Our method can correctly estimate the true spectrum identified by Monte Carlo sample variance. Recalling the common assumptions highlighted in Section II-B, we

[4] The constant $\frac{3}{8}$ of the Anscombe VST in [1] is here replaced by $\frac{7}{8}$ as we consider the heteroskedastic Gaussian instead of the actual Poisson distribution.

[5] It is easy to compute $\mathrm{var}\{f(z)|y\}$ from the first and second moments of $f(z)$ via trapezoidal integration.

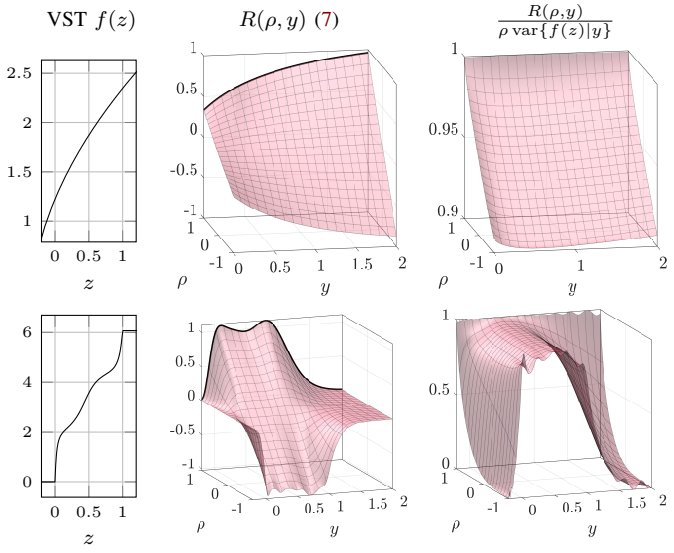

Fig. 2. Top and bottom rows correspond to the two examples of f and σ described in Section VI. From left to right: plot of the VST f; the effect of f on the covariance ρ represented by $R(\rho, y)$ (7) (pink mesh) and $\operatorname{var}\{f(z)|y\}$ (black solid curve at $\rho = 1$); the ratio between $R(\rho, y)$ and $\rho \operatorname{var}\{f(z)|y\}$. The fitting of f is computed with polynomials of degree $N = 46$ for both examples.

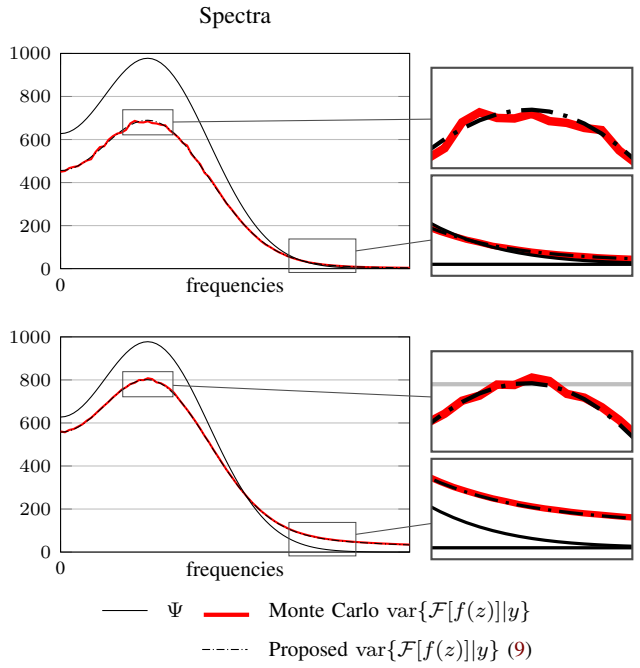

Fig. 3. Top and bottom rows for f and σ like in Fig. 2, here with noise PSD $\Psi = \operatorname{var}\{\mathcal{F}[\eta]\}$ resulted from an autocovariance function $\rho_\Psi(\xi) = \cos(\xi/2)e^{-\xi^2/18}$ on a 1D domain Ω with cardinality $|\Omega| = 257$; we consider $y = 1$ and $y = 0$ for the first and second example, respectively. For each example, the spectrum $\operatorname{var}\{\mathcal{F}[f(z)]|y\}$ of $f(z)$ computed by our proposed method (dash-dotted black line) matches closely the spectrum estimated as the sample variance of 10 000 Monte Carlo realizations of $\mathcal{F}[f(z)]\,|y$ (thick red line). Large approximation errors would result if one were instead to follow the common assumptions highlighted in Section II-B and use Ψ (solid black line) as the spectrum of $f(z)$. The PSDs are plotted only for the positive horizontal axis, as they are symmetric with respect to the vertical axis.

notice that in fact Ψ is far from approximating $\operatorname{var}\{\mathcal{F}[f(z)]|y\}$ for small y due to imperfect stabilization, thus showing that such assumptions are failing.

Fig. 3 also reports the results for the example with clipped data. Similar considerations to the first example hold here too. Additionally, we can notice[6] that in low frequencies Ψ is higher than $\operatorname{var}\{\mathcal{F}[f(z)]|y\}$, and in high frequencies Ψ is noticeably lower. Therefore, as mentioned in Section III, this is a case where the VST distorts the noise correlation to the extent that there is no $\lambda \in \mathbb{R}^+$ that can give $\operatorname{var}\{\mathcal{F}[f(z)]|y\} \approx \lambda\Psi$.

VII. Discussion and Conclusion

We propose a method that allows for modeling and efficient computation of the spectra of variance-stabilized data corrupted by correlated noise. We compute the spectra after stabilization from the autocovariance function of local polynomial approximations fitting the VST. Our model provides a higher accuracy than the spectrum model based on common simplifying assumptions. For the sake of verification, in this paper we used a large polynomial degree N, whereas in practical applications we expect lower degrees to suffice. Similar considerations apply to n too. Moreover, in cases like those illustrated in Fig. 2, the covariances vary smoothly, and it may be sufficient to compute the covariances for a coarse set of values of y, and resort to interpolation of either the PSDs or the covariances, depending on the use case.

Although in Section II-A we consider data subject to a signal-independent noise PSD, our proposed method is entirely based on conditional distributions and conditional PSDs, and therefore it can also work for data subject to noise PSD that are conditional on y and it returns PSDs of stabilized data that too are conditional on y. Typically, these conditional PSDs are indeed different for different y, as the polynomial coefficients $\boldsymbol{\alpha}^\star(y)$ (3) vary with y. Furthermore, although the proposed method was developed for VSTs, it can also be adopted to compute the spectra of data transformed by generic nonlinear univariate mappings, including parametric mappings that depend on y.

An apparent limitation of the method is the requirement of the noise η on $z|y$ to be Gaussian. However, as shown in Section VI, this limitation can be bypassed by expressing the non-Gaussian signal $\tilde{z}$ as originating from a latent Gaussian model z (1) through a transformation g, which is then composed with the VST and analyzed jointly. The general form of this transformation is $g_y(z) = F_y^{-1}(\Phi(\frac{z-y}{\sigma(y)}))$, where F_y^{-1} is the generalized inverse of the conditional cumulative distribution function (CDF) of $\tilde{z}|y$, and Φ is the CDF of the standard Gaussian distribution. In some particular cases, such as the clipping in Section VI or the noncentral chi-squared with one of freedom, this g_y can even be made independent of y. Therefore, the method is applicable also to many non-Gaussian noise distributions of significant practical relevance.

[6] Similar distortions have been observed with other nonlinear functions too [18].

REFERENCES

[1] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3-4, pp. 246–254, 1948.

[2] B. Efron, "Transformation theory: How normal is a family of distributions?" *The Annals of Statistics*, vol. 10, no. 2, pp. 323–339, 1982.

[3] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009.

[4] L. R. Borges, M. A. C. Brochi, Z. Xu, A. Foi, M. A. C. Vieira, and P. M. Azevedo-Marques, "Noise modeling and variance stabilization of a computed radiography (CR) mammography system subject to fixed-pattern noise," *Physics in Medicine & Biology*, vol. 65, no. 22, p. 225035, Nov. 2020.

[5] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support." *Astronomy and Astrophysics Supplement*, vol. 112, p. 179, 1995.

[6] P. R. Prucnal and E. L. Goldstein, "Exact variance-stabilizing transformations for image-signal-dependent rayleigh and other weibull noise sources," *Applied optics*, vol. 26, no. 6, pp. 1038–1041, 1987.

[7] M. Mäkitalo, A. Foi, D. Fevralev, and V. Lukin, "Denoising of single-look SAR images based on variance stabilization and nonlocal filters," in *2010 International Conference on Mathematical Methods in Electromagnetic Theory*. Kyiv, Ukraine: IEEE, Sep. 2010, pp. 1–4.

[8] A. Foi, "Noise estimation and removal in MR imaging: The variance-stabilization approach," in *2011 IEEE International symposium on biomedical imaging: from nano to macro*. IEEE, 2011, pp. 1809–1814.

[9] Y. Mäkinen, S. Marchesini, and A. Foi, "Ring artifact and Poisson noise attenuation via volumetric multiscale nonlocal collaborative filtering of spatially correlated noise," *Synchrotron Radiation*, vol. 29, no. 3, pp. 829–842, 2022.

[10] S. Herbreteau and M. Unser, "Self-calibrated variance-stabilizing transformations for real-world image denoising," *arXiv preprint arXiv:2407.17399*, 2024.

[11] Y. Hu, H. Xu, X. Zhu, and H. N. Wake, "V-DDPM: MRI Rician noise removal model based on VST and DDPM," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2250–2254.

[12] Y. Mäkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Transactions on Image Processing*, vol. 29, pp. 8339–8354, 2020.

[13] L. Azzari and A. Foi, "Variance stabilization in Poisson image deblurring," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. Melbourne, Australia: IEEE, Apr. 2017, pp. 728–731.

[14] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1/2, p. 134, Nov. 1918.

[15] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[16] L. Azzari, L. R. Borges, and A. Foi, "Modeling and estimation of signal-dependent and correlated noise," in *Denoising of Photographic Images and Video*, M. Bertalmío, Ed. Cham: Springer International Publishing, 2018, pp. 1–36, series Title: Advances in Computer Vision and Pattern Recognition.

[17] A. Foi, "Optimization of variance-stabilizing transformations," *Preprint, 2009b*, 2009, https://webpages.tuni.fi/foi/papers/Foi-OptimizationVST-preprint.pdf.

[18] J. Van Vleck and D. Middleton, "The spectrum of clipped noise," *Proceedings of the IEEE*, vol. 54, no. 1, pp. 2–19, 1966.