# Contents

# Modeling and Estimation of Signal-Dependent and Correlated Noise

Lucio Azzari, Lucas Rodrigues Borges, Alessandro Foi

**Abstract** The additive white Gaussian noise (AWGN) model is ubiquitous in signal processing. This model is often justified by central-limit theorem (CLT) arguments. However, whereas the CLT may support a Gaussian distribution for the random errors, it does not provide any justification for the assumed additivity and whiteness. As a matter of fact, data acquired in real applications can seldom be described with good approximation by the AWGN model, especially because errors are typically correlated and not additive. Failure to model accurately the noise leads to inaccurate analysis, ineffective filtering, and distortion or even failure in the estimation. This chapter provides an introduction to both signal-dependent and correlated noise and to the relevant models and basic methods for the analysis and estimation of these types of noise. Generic one-parameter families of distributions are used as the essential mathematical setting for the observed signals. The distribution families covered as leading examples include Poisson, mixed Poisson-Gaussian, various forms of signal-dependent Gaussian noise (including multiplicative families and approximations of the Poisson family), as well as doubly censored heteroskedastic Gaussian distributions. We also consider various forms of noise correlation, encompassing pixel and readout cross-talk, fixed-pattern noise, column/row noise, etc., as well as related issues like photo-response and gain nonuniformity. The introduced models and methods are applicable to several important imaging scenarios and technologies, such as raw data from digital camera sensors, various types of radiation imaging relevant to security and to biomedical imaging.

———————————

Lucio Azzari

Tampere University of Technology, Korkeakoulunkatu 1, 33720 Tampere, Finland, e-mail: lucio.azzari@tut.fi

Lucas Rodrigues Borges

São Carlos School of Engineering, University of São Paulo, 400 Trabalhador São-Carlense Avenue, São Carlos 13566-590, Brazil, e-mail: lucas.rodrigues.borges@usp.br

Alessandro Foi

Tampere University of Technology, Korkeakoulunkatu 1, 33720 Tampere, Finland, e-mail: alessandro.foi@tut.fi

# 1 Introduction: Acquisition Devices and Noise Sources

A digital image is generated by converting the light coming from a natural scene to numerical pixel values. In particular, a typical camera performs this conversion using a semiconducting array of sensing elements positioned after an aperture: when the shutter opens, the light from the scene goes through the lenses and the aperture, finally colliding with the sensor array. Each element in the array converts the energy of the incident light beam to electric charges that are successively accumulated in an electric potential. The electric potentials are then converted to digital values, and finally stored collectively as a raw image, whose pixel values are ideally proportional to the intensity of the light that shone onto the corresponding sensing elements.

The most common digital camera sensors are Charge Coupled Semiconductor Devices (CCD) and Complementary Metal-Oxide Semiconductor (CMOS). While CCD used to be the most common technology, nowadays CMOS sensors dominate the market, being the preferred capture technology for smartphones and digital cameras. The main difference between the two is that, while in CCD arrays the charge of a row of sensors is transported via the same circuit, sharing also the same amplifier, CMOS arrays are based on the Active Pixel Sensor (ASP) technology, for which every single sensor is treated independently, having a unique transport line [Jähne, 2004].

To get a basic understanding of the nature of the noise in imaging sensors, let us consider the acquisition of a still scene; although the average incident energy over a relatively long period of time might be virtually constant, the amount of photons incident on the camera sensors during the exposure fluctuates in time. Furthermore, not all the incident photons are converted to electric charge. This whole phenomenon is known as shot noise, and it is well modeled by the family of Poisson distributions [Mandel, 1959]. An important feature of this type of noise is that it is *signal-dependent*, in the sense that the electric charge fluctuates in time with a variance that is proportional to the photon flux. Thus, different parts of a captured scene are subject to different noise strengths with the stronger noise affecting the brighter content.

Another relevant source of noise is the so-called thermal noise. Thermal noise is generated by thermal agitation [Nyquist, 1928; Johnson, 1928] and is due to the fact that at any given temperature (except absolute zero), conductors have a probability to emit charges due to heat, even when there is no electric potential to stimulate them. This results in a background current, present also in the absence of input signals (dark current) [Jähne, 2004], which alters the measurements of the sensors. The inevitable fluctuations of this current are thus modeled as noise. By definition, this type of noise is proportional to the working temperature of the device, and it can therefore be reduced by decreasing the temperature of sensor. In high-end devices for scientific applications (e.g., optical astronomy), this is achieved by means of a thermoelectric cooler such as a Peltier heat pump; however, in most consumer applications, sensor cooling is not feasible and thermal noise, suitably modeled by a Gaussian distribution, becomes a significant component of the measurement er-

rors. Particularly when capturing scenes in low-light conditions, thermal noise can dominate the overall noise.

Furthermore, there is a possibility that the charges accumulated by neighboring sensor elements coupled with each other introducing correlation between measured quantities. In other words, the charge accumulated by a sensor element is influenced not only by the number of incident photons, but also by the charges accumulated by the surrounding sensor elements. Analogously, during the readout phase, when the device reads and transports the charges from the sensor elements, there could be some electrical coupling of the quantities. This introduces an error in the acquisition process that is commonly referred to as *cross-talk*, and it is usually well modeled by the adoption of correlated noise, in which the measurement error for a pixel is influenced also by the surrounding errors.

Finally, the electric potential is often amplified (analog gain) before being converted to a digital value by an analog-to-digital converter. This analog amplification may introduce further noise and, because the digital values are discrete with a certain bit depth, we eventually encounter also quantization noise, which is sometimes approximated by uniformly distributed errors over one quantization step, or as a generic additive noise with comparable variance (i.e., one-twelfth of the quantization step, as per basic properties of the uniform distribution).

## 2 Additive White Gaussian Noise

As highlighted above, a signal acquired by a digital device is affected by noise from several sources. It is often difficult to separate and treat each noise source individually, as this requires in-depth knowledge of the device and direct access to some of its inner components; therefore, the various sources are conventionally grouped together and addressed as a single noise process. This procedure is encouraged by the central-limit theorem (CLT) [Tijms, 2007; Papoulis and Pillai, 2002]: for a set of $N$ independent random variables $X_1, \ldots, X_N$, with respective means $\mu_1, \ldots, \mu_N$ and standard deviations $\sigma_1, \ldots, \sigma_N$, as $N \to \infty$ we have

$$\frac{1}{s} \sum_{i=1}^{N} (X_i - \mu_i) \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{with} \qquad s = \sqrt{\sum_{i=1}^{N} \sigma_i^2}, \tag{1}$$

where $\xrightarrow{d}$ denotes the convergence in distribution, and $\mathcal{N}(0,1)$ indicates a Gaussian (also called normal) random variable with mean and variance equal, respectively, to the first and second arguments within parenthesis (in this case 0 and 1). In other words, we can represent the sum of various random noise sources as a Gaussian random variable, irrespective of the noise distribution of the individual sources. The CLT establishes the importance of the Gaussian distribution in modeling complex physical processes.

The Gaussian noise is further commonly assumed additive, zero-mean, independent, and identically distributed (i.i.d.); under these extra assumptions, a captured image $z$ is modeled as

$$z(x) = y(x) + \eta(x), \tag{2}$$

where $y$ is the underlying deterministic noise-free image, $x \in \Omega \subset \mathbb{Z}^2$ is the pixel coordinate, and $\eta(\cdot) \sim \mathcal{N}(0, \sigma^2)$ is the zero-mean Gaussian random variable with variance $\sigma^2$. Each coordinate $x$ results in an independent (hence different) realization of the random variable $\eta(x)$, which collectively for all $x \in \Omega$ yields the additive white Gaussian noise (AWGN) field corrupting $y$. The term *white* is inspired by spectroscopy: like white light dispersed through a prism reveals components for every frequency in the visible spectrum (from 400 THz of red to 789 THz of violet), Fourier analysis of white noise reveals components for every frequency within the Fourier spectrum. Specifically, when working on a 2D rectangular image domain of $N_1 \times N_2$ pixels, i.e., $x = (x_1, x_2) \in \Omega = [0, \ldots, N_1 - 1] \times [0, \ldots, N_2 - 1]$, a generic Fourier coefficient of $\eta$ can be written as

$$\mathscr{F}[\eta](\xi_1, \xi_2) = \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} e^{-2\pi i \left(\xi_1 \frac{x_1}{N_1} + \xi_2 \frac{x_2}{N_2}\right)} \eta(x_1, x_2), \tag{3}$$

where $\mathscr{F}$ denotes the Fourier transform and $\xi_1$, $\xi_2$ are spatial frequencies. The noise power spectrum (also called power spectral density, PSD) corresponds to the variance of $\mathscr{F}[\eta]$, which can be computed as

$$\mathrm{var}\{\mathscr{F}[\eta](\xi_1, \xi_2)\} = \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} \mathrm{var}\left\{ e^{-2\pi i \left(\xi_1 \frac{x_1}{N_1} + \xi_2 \frac{x_2}{N_2}\right)} \eta(x_1, x_2) \right\} = \tag{4}$$

$$= \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} \left| e^{-2\pi i \left(\xi_1 \frac{x_1}{N_1} + \xi_2 \frac{x_2}{N_2}\right)} \right|^2 \mathrm{var}\{\eta(x_1, x_2)\} = \tag{5}$$

$$= \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} \mathrm{var}\{\eta(x_1, x_2)\} = \tag{6}$$

$$= \sum_{x_2=0}^{N_2-1} \sum_{x_1=0}^{N_1-1} \sigma^2 = N_1 N_2 \sigma^2, \tag{7}$$

i.e., the power (variance) of the noise is constant in the Fourier domain and directly proportional to the variance in the pixel domain. We can say that the Fourier spectrum of white noise is *flat*. Equalities (4)-(7) leverage few basic properties: (1) $\eta$ is independently distributed; (2) multiplication of $\eta$ by a deterministic function (i.e., the complex exponential) does not affect the independence; hence (3) the Fourier coefficient $\mathscr{F}[\eta](\xi_1, \xi_2)$ (3) is simply a sum of independent random variables; (4) the variance of the sum of independent random variables is the sum of their variances (4); (5) multiplication of a random variable by a deterministic factor scales the variance by the squared modulus of the factor (5); (6) complex exponentials are always on the unit circle of the complex plane, i.e., they have unit modulus

(6); (7) $\eta$ is identically distributed with constant variance $\sigma^2$ (7). We can see that of the i.i.d. hypothesis, the independence alone is sufficient to reach (6) and that having identical distributions (hence, identical variance $\sigma^2$ for every $x \in \Omega$) is used only for obtaining (7). Already (6) shows that the Fourier spectrum of the noise is flat, since $\text{var}\{\mathscr{F}[\eta](\xi_1, \xi_2)\}$ no longer depends on $\xi_1$ or $\xi_2$. Indeed, we can have white noise with a flat Fourier power spectrum for models different from the AWGN (2).

At this point, it is important to emphasize that although theoretically the CLT justifies using a Gaussian distribution for modeling the measurement random errors, it does not provide arguments supporting the additional assumptions of the AWGN model, namely, that the errors are independent and identically distributed over the image. Indeed, in (1), we can observe that the term on the left-hand side is essentially a standardization of the errors, which thus explicitly depends on the means $\mu_i$ and variances $\sigma_i^2$ of the individual contributors. In general, we have that each noisy pixel $z(x)$ results from its own sequence $X_i(x)$, $i = 1, \ldots, N$, where the means and the variances of these contributors can be different at different pixels; in other words, the errors may not be identically distributed and even when a Gaussian model (as per the CLT) may be appropriate, then mean and (most often) the variance of such Gaussian errors may change from pixel to pixel. Furthermore, contributors of different pixels can be subject to a mutual interaction, possibly resulting in a statistical dependence between the measurement errors at different pixels; in other words, the errors may not be independently distributed over the image.

It is clear from these premises and from the summary in Sect. 1 that the AWGN model is not suitable for modeling the above measurement errors, first because it inherently uses a single distribution, and second because it assumes independent errors.

In Sect. 4, we begin from generalizing the observation model to accommodate a multiplicity of distributions. Specifically, we adopt the formalism of one-parameter families of distributions, where the observation at each pixel follows a specific distribution that depends on a known or unknown univariate parameter. Further, in Sect. 6.1, we address the issue of correlation in the errors.

## 3 Raw Image Dataset

Throughout this chapter, we use real sensor raw data to validate the presented models and methods. As leading example, we use a dataset consisting of $M = 30$ raw images of the same still scene acquired repeatedly in a short time interval (at a rate of about 1 frame/s) under identical capture settings. We identify the individual images in the set as $\tilde{z}^{(m)}$, $m = 1, \ldots, 30$, while $\tilde{z}$ denotes a generic such image; the reason for using the tilde decoration here will become clear in the further sections. The images have been captured by a Samsung S5K2L2 CMOS ISOCELL sensor with a $1.4\mu m$ pixel size at ISO 1250; this type of sensor can be found in modern mobile devices such as the Samsung Galaxy S8 smartphone. Raw images from this sensor are stored in 10-bit format, which we normalize to the range $[0, 1]$ by dividing the raw integer
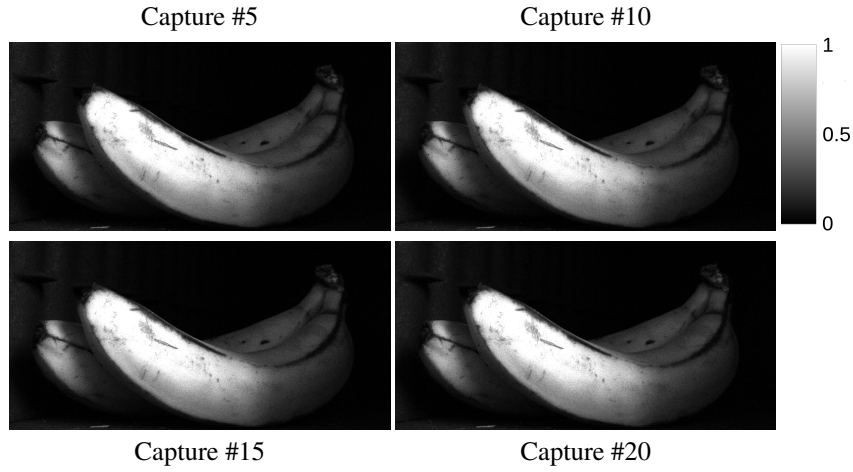
Capture #5                                Capture #10



Capture #15                               Capture #20

Fig. 1: Examples from the dataset of 30 raw images captured under identical settings with a Samsung S5K2L2 CMOS ISOCELL sensor at ISO 1250.

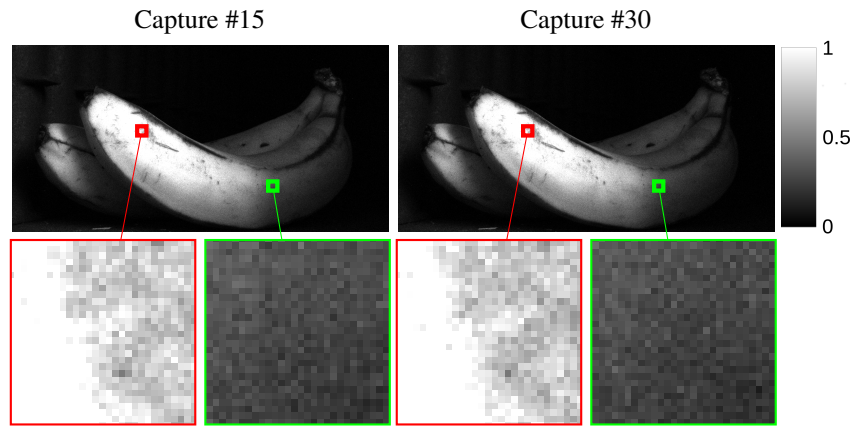Capture #15                               Capture #30



Fig. 2: Two raw images of the same scene, captured under identical settings. A total of 30 images like these were captured.

values by $2^{10} - 1$. Four images from the dataset are reported in Fig. 1: the scene features a dark background with two bananas in the foreground. The presence of both dark and bright regions makes the dataset suitable for validating the noise models described in the following sections. Even though these sensors are typically used in conjunction with a color filter array such as the RGGB Bayer filter mosaic, for the sake of simplicity of presentation we consider only single-channel monochromatic (green) acquisition.
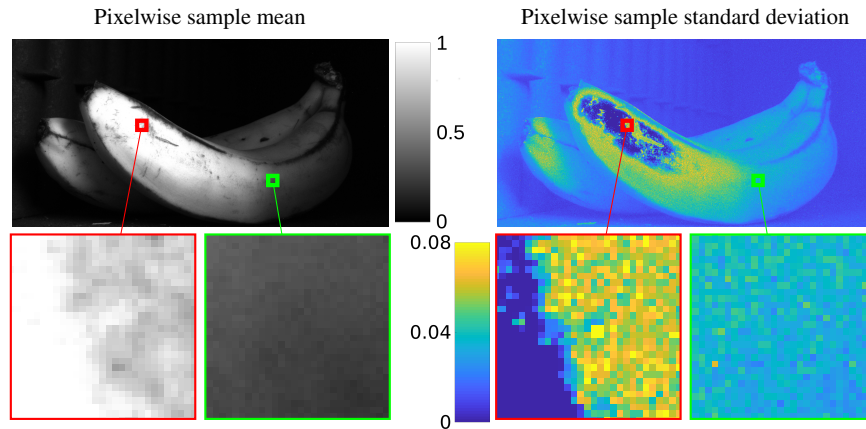
Fig. 3: Pixelwise sample mean (left) and sample standard deviation (right) across all the captured images of the dataset.
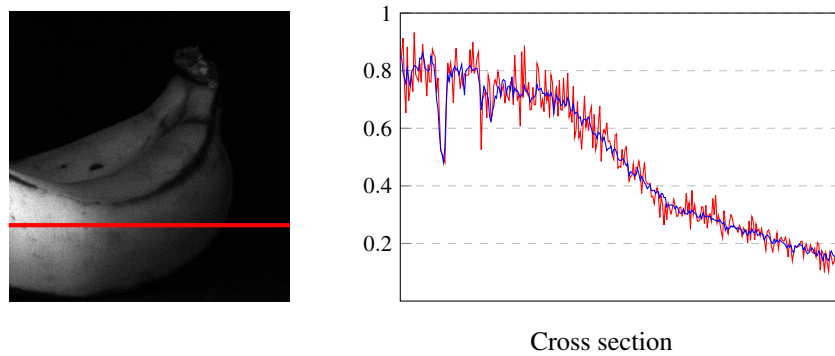


Cross section

Fig. 4: Left: detail from the dataset with highlighted cross section. Right: cross section (red line) plotted against its expectation $\mathrm{E}\widehat{\{\tilde{z}(x)\}}$ (in blue). Note how the noise is *signal-dependent*, with different variance at each pixel $x$ depending on the value of the underlying expectation.

In the top row of Fig. 2, we show two images of the dataset. By inspecting the enlarged fragments, we can observe how the two images are practically equal except for the individual noise realizations. Indeed, we can formally define the image noise as the difference between the captured images and their mathematical expectation, i.e., the average of infinitely many images like those captured in the dataset, where the latter, denoted by $\tilde{y}$, can be treated as the ideal noise-free image. Since the dataset contains only finitely many images ($M = 30$), we can approximate the mathematical expectation by the pixelwise sample average,

$$\tilde{y}(x) = \mathrm{E}\{\tilde{z}(x)\} \approx \frac{1}{M} \sum_{m=1}^{M} \tilde{z}^{(m)}(x) = \widehat{\mathrm{E}\{\tilde{z}(x)\}}, \tag{8}$$

resulting in the image in the left-hand side of Fig. 3. Note in the enlarged fragments how the noise is virtually removed everywhere.

The cross section shown in Fig. 4 introduces the reader to a main feature that is shared by most of the noise types under consideration: noise affecting the bright parts of the image is significantly stronger (i.e., larger errors) compared to the noise affecting dark regions. This can be quantified as the standard deviation of the noise at each pixel (again, computed over infinitely many such captured images), which we can approximate by the pixelwise sample standard deviation over the finite dataset,

$$\mathrm{std}\{\tilde{z}(x)\} \approx \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} \left(\tilde{z}^{(m)}(x) - \frac{1}{M} \sum_{l=1}^{M} \tilde{z}^{(l)}(x)\right)^2} = \widehat{\mathrm{std}\{\tilde{z}(x)\}}, \tag{9}$$

which is shown in the image in the right-hand side of Fig. 3. To provide a visual exploration of the relation between the expectation and the standard deviation of the noisy raw pixels, Fig. 5 shows a scatterplot where each red dot represents a $\left(\widehat{\mathrm{E}\{\tilde{z}(x)\}}, \widehat{\mathrm{std}\{\tilde{z}(x)\}}\right)$ pair for $x \in \Omega$. The scatterplot can be interpreted as a cloud of points about an unknown smooth curve that describes the noise standard deviation as function of the signal expectation. However, this interpretation is admissible only if the dispersion of the scatterpoints is compatible with the existence of such curve. Indeed, leveraging the CLT and the first-order Taylor expansion of the square root at $\mathrm{var}\{\tilde{z}(x)\}$, the distributions of $\widehat{\mathrm{E}\{\tilde{z}(x)\}}$ and $\widehat{\mathrm{std}\{\tilde{z}(x)\}}$ can be approximated for large $M$ as

$$\widehat{\mathrm{E}\{\tilde{z}(x)\}} \sim \mathcal{N}\left(\mathrm{E}\{\tilde{z}(x)\}, \frac{1}{M}\mathrm{var}\{\tilde{z}(x)\}\right), \tag{10}$$

$$\widehat{\mathrm{std}\{\tilde{z}(x)\}} \sim \mathcal{N}\left(\mathrm{std}\{\tilde{z}(x)\}, \frac{2+\kappa}{4M}\mathrm{var}\{\tilde{z}(x)\}\right), \tag{11}$$

where $\kappa$ is the excess kurtosis of $\tilde{z}(x)$. Taking into account the sample histograms plotted in Fig. 6, the distributions (10)-(11) fully explain the dispersion visible in the scatterplot and suggest a functional relation between $\mathrm{E}\{\tilde{z}(x)\}$ and $\mathrm{std}\{\tilde{z}(x)\}$, which may be obtained, e.g., by processing the scatterplot with a smoother. The histograms also illustrate that $\tilde{z}$ is not identically distributed and that its distribution varies from pixel to pixel according to the expectation.

Another important feature that can be ascertained from Fig. 2 and Fig. 3 is the fact that the brightest areas of the image *saturate* to white. This phenomenon is commonly referred to as *clipping* and results, in particular, in the drop of sample standard deviation that can be observed in Fig. 3 and Fig. 5, and in the lack of Gaussianity in some of the histograms in Fig. 6.
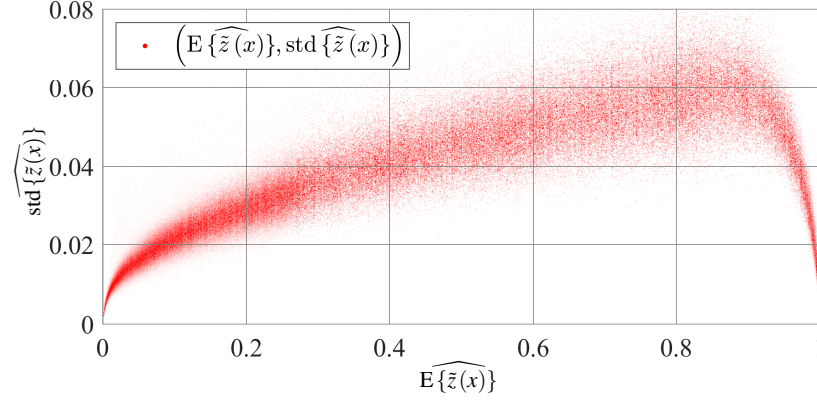
Fig. 5: Scatterplot of the pairs $\left( \mathrm{E}\,\widehat{\{\tilde{z}(x)\}}, \mathrm{std}\,\widehat{\{\tilde{z}(x)\}} \right)$ drawn as red dots. The dispersion visible in the scatterplot is described by the distributions (10)-(11) of the estimated pairs.
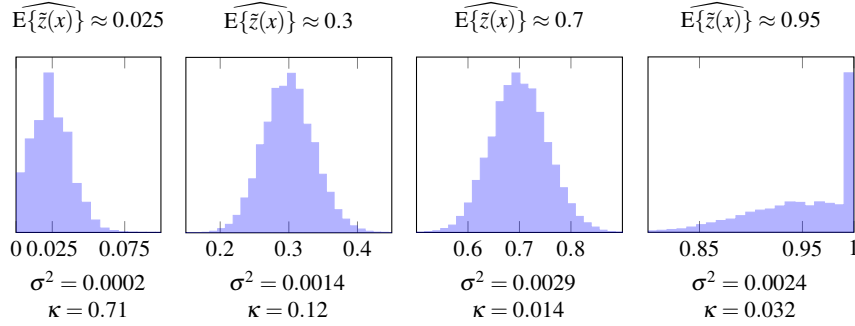


Fig. 6: Histograms of the pixels $\tilde{z}(x)$ from the dataset with pointwise sample mean $\mathrm{E}\,\widehat{\{\tilde{z}(x)\}} \approx [0.025, 0.3, 0.7, 0.95]$. Below each histogram we report its variance $\sigma^2$ and excess kurtosis $\kappa$.

Overall, the above analysis indicates that the noise affecting the dataset images is *signal-dependent* in the sense that its characteristics at each pixel $x$ depend on the value of the underlying noise-free image, i.e., on $\tilde{y}(x) = \mathrm{E}\,\{\tilde{z}(x)\}$.

In the next section, we will derive, step-by-step, a simple yet effective mathematical model that accurately describes the behavior of the points of the scatterplot in Fig. 5, that provides a direct functional relation between $\mathrm{E}\,\{\tilde{z}(x)\}$ and $\mathrm{std}\,\{\tilde{z}(x)\}$ and that explains the shape and moments of the histograms as a function of $\mathrm{E}\,\{\tilde{z}(x)\}$.
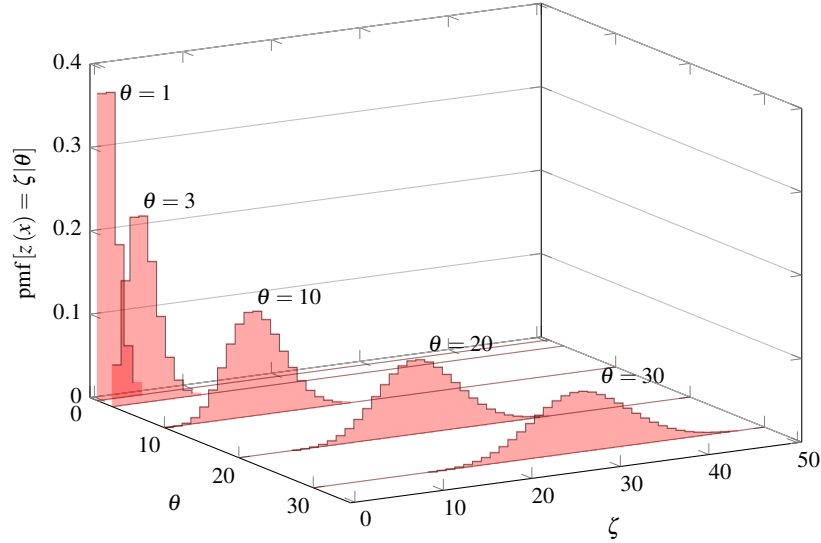
Fig. 7: Illustration of the family of Poisson distributions (12). We show the distributions with mean $y(x) = \theta = 1, 3, 10, 20, 30$.

,

# 4 One-Parameter Families of Distributions

A one-parameter family of distributions $\{\mathscr{D}_\theta\}_{\theta \in \Theta}$ is a collection of distributions, each of which is identified by the value of a univariate parameter $\theta \in \Theta \subseteq \mathbb{R}$.

Let $z$ be a random variable distributed according to a one-parameter family of distributions $\{\mathscr{D}_\theta\}_{\theta \in \Theta}$: this means that for each individual $\theta \in \Theta$, the conditional distribution of $z$ given $\theta$ is $\mathscr{D}_\theta$, i.e., $z|\theta \sim \mathscr{D}_\theta$. Hence, the conditional expectation and the conditional standard deviation of $z$ given $\theta$, i.e., $\mathrm{E}\{z|\theta\}$ and $\mathrm{std}\{z|\theta\}$, are two functions of $\theta$.

In the following sections, we cover some of the most important one-parameter families of distributions for modeling noise of digital imaging sensors, describing them in detail through the corresponding probability density functions (PDFs) or probability mass functions (PMFs), and their mean and variances.

## 4.1 Poisson Noise and Poisson Family of Distributions

The simplest way to model an image captured by a photodetector array is to represent it as a realization of independent *Poisson* random variables. In particular, the captured image $z$ is modeled as

$$z(x) \sim \mathscr{P}(y(x)),$$

$$\mathrm{pmf}[z(x) = \zeta \mid y(x)] = \begin{cases} \frac{y(x)^{\zeta} e^{-y(x)}}{\zeta!} & \zeta \in \mathbb{N} \cup \{0\} \\ 0 & \text{elsewhere,} \end{cases} \tag{12}$$

where $y \geq 0$ is the noise-free image, which can be thought as a proxy for the photon flux, and the symbol $\mathscr{P}$ denotes the Poisson family of distributions. This is a one-parameter family of distributions with parameter $\theta = y(x)$, which coincides with the mean and variance of the conditional distributions:

$$\mathrm{E}\{z(x) \mid y(x)\} = \mathrm{var}\{z(x) \mid y(x)\} = y(x). \tag{13}$$

Figure 7 shows examples of distributions from the Poisson family.

*Poisson noise* can be formally defined as $z - y$. It is clear from (13) that the mean of Poisson noise is zero, i.e., $\mathrm{E}\{z(x) - y(x) \mid y(x)\} = 0$, and its variance is $y$, i.e., $\mathrm{var}\{z(x) - y(x) \mid y(x)\} = \mathrm{var}\{z(x) \mid y(x)\} = y(x)$. From (13), we can also observe an important property of Poisson noise: since the variance is equal to the mean of the signal, there is a square root relation between mean and standard deviation. This implies that the signal-to-noise ratio SNR

$$\mathrm{SNR} = \frac{\mathrm{E}\{z(x) \mid y(x)\}}{\sqrt{\mathrm{var}\{z(x) \mid y(x)\}}} = \frac{y(x)}{\sqrt{y(x)}} = \sqrt{y(x)} \tag{14}$$

increases when the intensity of the noise-free signal $y$ increases and converges to zero when $y$ approaches zero. This means that Poisson images captured at lower intensities, even though in absolute terms feature a lower variance, they are in practice noisier relative to their mean intensity, and when $y < 1$ they are effectively dominated by noise. Such conditions correspond to what is commonly termed *photon-limited imaging*, which is one of the most challenging imaging scenarios, requiring binning or special denoising procedures (see, e.g., [Azzari and Foi, 2016]). Although infrequent in the context of consumer imaging and photography, photon-limited imaging is an increasingly important scenario in scientific imaging, particularly in astronomical imaging, fluorescence microscopy, and low-dose radiation imaging for medicinal diagnostics.

### 4.2 Scaled Poisson Distribution Family

In many cases, it is convenient to use the so-called *scaled Poisson* distributions, where a positive scaling parameter controls the noise variance relative to the signal mean. Such scaling factor is commonly used to model the *quantum efficiency* of the imaging sensor, i.e., the ratio between the (average) number of converted electrons to the number of incident photons.

A scaled Poisson distribution with scale parameter $a > 0$ and mean $y(x) \geq 0$ is of the form

$$a^{-1}z(x) \sim \mathscr{P}\left(a^{-1}y(x)\right), \tag{15}$$

$$\mathrm{pmf}\left[z(x) = \zeta \mid y(x)\right] = \begin{cases} \frac{(y(x)/a)^{\zeta/a}e^{-y(x)/a}}{(\zeta/a)!} & \zeta \in \{0, a, 2a, 3a, \dots\} \\ 0 & \text{elsewhere.} \end{cases} \tag{16}$$

The resulting family of distributions can be parametrized by $\theta = y(x)$. The mean and variance of the scaled Poisson distributions are

$$\begin{aligned} \mathrm{E}\left\{z(x) \mid y(x)\right\} &= y(x), \\ \mathrm{var}\left\{z(x) \mid y(x)\right\} &= ay(x). \end{aligned} \tag{17}$$

Observe that the parameter $a$ scales only the variance but does not affect the expectation; hence, it controls the relative strength of the noise, and in particular we have

$$\mathrm{SNR} = \frac{\mathrm{E}\left\{z(x) \mid y(x)\right\}}{\sqrt{\mathrm{var}\left\{z(x) \mid y(x)\right\}}} = \sqrt{\frac{y(x)}{a}}. \tag{18}$$

### 4.3 Poisson-Gaussian Noise

The Poisson-Gaussian noise model is given by the sum of two independent sources of noise: Poisson or scaled Poisson, whose variance is signal-dependent (and proportional to the signal mean) and Gaussian, whose variance is signal-independent. Its formal model is

$$z(x) = ap(x) + n(x), \tag{19}$$

where

$$p(x) \sim \mathscr{P}\left(a^{-1}y(x)\right) \qquad \text{and} \qquad n(x) \sim \mathscr{N}(0, b), \tag{20}$$

and the constants $a > 0$ and $b \geq 0$ are, respectively, the scaling factor for the scaled Poisson addend $ap$ and the variance of the Gaussian addend $n$. We have

$$\mathrm{pdf}\left[z(x) \mid y(x)\right](\zeta) = \sum_{k=0}^{+\infty} \frac{(y(x)/a)^k e^{-y(x)/a}}{k!} \times \frac{1}{\sqrt{2\pi b}} e^{-\frac{(\zeta - ka)^2}{2b}}, \tag{21}$$

which also corresponds to a one-parameter family of distributions with parameter $\theta = y(x)$. The conditional mean and variance of $z$ are

$$\begin{aligned} \mathrm{E}\left\{z(x) \mid y(x)\right\} &= y(x) \\ \mathrm{var}\left\{z(x) \mid y(x)\right\} &= ay(x) + b. \end{aligned} \tag{22}$$

The *Poisson-Gaussian noise* is formally defined as $z(x) - y(x)$, and by (22) it has zero-mean and affine variance on $y$.
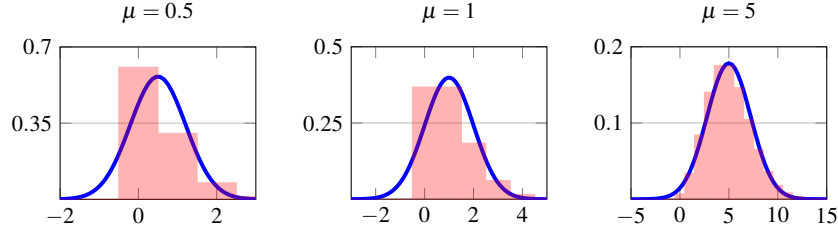
The SNR is calculated as

Fig. 8: Gaussian approximation of the Poisson distribution. From left to right we draw in red the discrete Poisson distributions $\mathscr{P}(\mu)$, and in blue their Gaussian approximations $\mathscr{N}(\mu,\mu)$ for three different mean values $\mu = [0.5, 1, 5]$. Note how the accuracy of the approximation improves as $\mu$ increases.

$$\text{SNR} = \frac{\text{E}\{z(x)\,|\,y(x)\}}{\sqrt{\text{var}\{z(x)\,|\,y(x)\}}} = \frac{y(x)}{\sqrt{ay(x)+b}}. \tag{23}$$

### 4.4 Gaussian Approximation of the Poisson Distribution

For large mean values, the Poisson distribution is well approximated by a Gaussian distribution with mean and variance equal to the mean of the Poisson random variable:

$$\mathscr{P}(\mu) \approx \mathscr{N}(\mu,\mu). \tag{24}$$

Here, we derive an intuitive proof based on the CLT and on the fact that the Poisson distributions are closed family with respect to summation of variables.

Let us consider two independent Poisson random variables $X_1$ and $X_2$, and their sum $Y = X_1 + X_2$. Assuming that $X_1$ and $X_2$ have means $\mu_{X_1}$ and $\mu_{X_2}$, respectively, the probability that the sum $X_1 + X_2$ takes a given value $\zeta \in \mathbb{N}$ is the sum of the probabilities that $X_1$ takes value $i \in \mathbb{N}$ and $0 \le i \le \zeta$ and $X_2$ takes value $\zeta - i$ (thus summing up to $\zeta$):

$$\text{pmf}[X_1 + X_2 = \zeta] = \sum_{i=0}^{\zeta} \text{pmf}[X_1 = i]\,\text{pmf}[X_2 = \zeta - i] = \sum_{i=0}^{\zeta} \frac{\mu_{X_1}^i e^{-\mu_{X_1}}}{i!} \frac{\mu_{X_2}^{\zeta-i} e^{-\mu_{X_2}}}{(\zeta - i)!}$$

$$= e^{-(\mu_{X_2}+\mu_{X_2})} \sum_{i=0}^{\zeta} \frac{\mu_{X_1}^i \mu_{X_2}^{\zeta-i}}{i!\,(\zeta - i)!} = e^{-(\mu_{X_1}+\mu_{X_2})} \frac{1}{\zeta!} \sum_{i=0}^{\zeta} \binom{\zeta}{i} \mu_{X_1}^i \mu_{X_2}^{\zeta-i}$$

$$= \frac{(\mu_{X_1}+\mu_{X_2})^{\zeta} e^{-(\mu_{X_1}+\mu_{X_2})}}{\zeta!},$$

$$\tag{25}$$

which shows that $Y = X_1 + X_2$ is a Poisson random variable with mean and variance $\mu_{X_1} + \mu_{X_2}$, i.e., $Y \sim \mathscr{P}(\mu_{X_1} + \mu_{X_2})$. Hence, any Poisson random variable with large
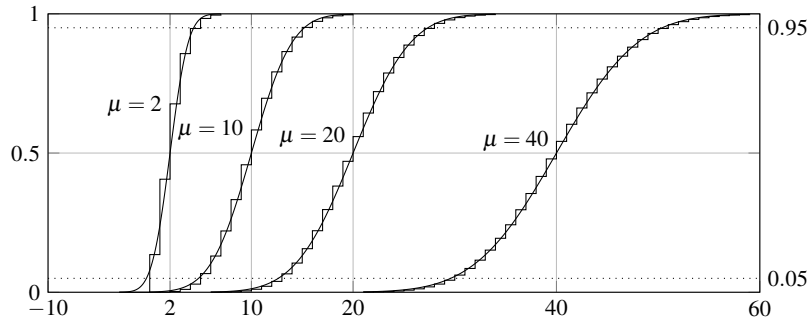
Fig. 9: Cumulative distribution functions of the Poisson $\mathscr{P}(\mu)$ and of the Gaussian $\mathscr{N}(\mu, \mu)$, $\mu = 2, 10, 20, 40$, showing convergence in distribution for large $\mu$.

enough mean can be expressed by a summation of many Poisson random variables with smaller means. Therefore, according to the CLT, as the mean value increases, the Poisson distribution converges in distribution to a Gaussian distribution with mean and variance equal to the mean of the Poisson random variable. Figure 8 shows in red three Poisson distributions with means $\mu = [0.5, 1, 5]$ overimposed to three Gaussian distributions, blue lines, with means and variances equal to $\mu$. We can observe that already for $\mu = 5$ the Gaussian provides a relatively good approximation of the Poisson distribution. This is further illustrated by the cumulative distribution functions shown in Fig. 9 for $\mu = 2, 10, 20, 40$. Most imaging applications deal with Poisson models well above such values and can thus leverage the approximation (24).

With the above approximation in mind, in many applications, it is possible to replace the family of Poisson distributions with a family of Gaussian distributions with nonconstant variance that depends on the signal expectation. This approximation in the modeling is appealing as it often results in simplification of analysis and processing operations such as noise estimation and denoising.

### 4.5 Signal-Dependent Heteroskedastic Gaussian Models

As a consequence of (24), we can approximate the Poisson-Gaussian model (19) by the sum of a deterministic signal $y(x)$ and two zero-mean Gaussian random variables, one with signal-independent variance $b$ and one with signal-dependent variance $ay(x)$. Since the sum of two zero-mean Gaussian random variables is still a zero-mean Gaussian random variable with variance equal to the sum of the variances, we have

$$z(x) = y(x) + \sigma(y(x))\xi(x), \tag{26}$$
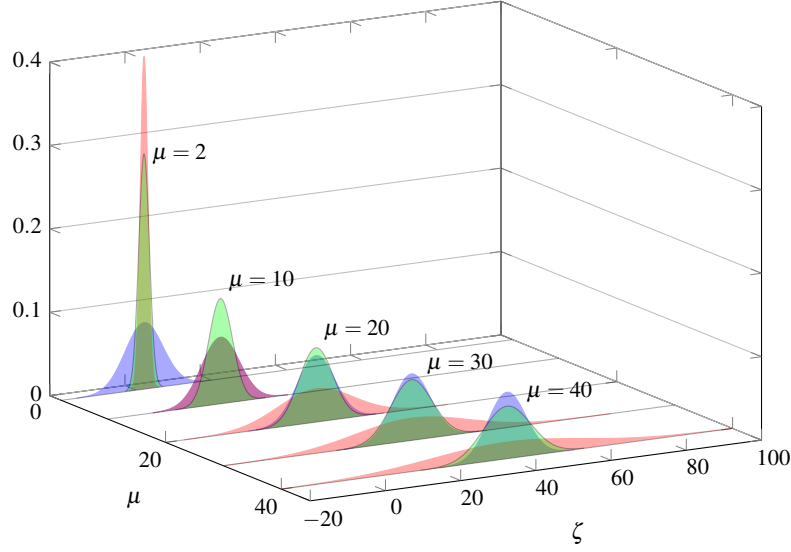
where $\xi(x) \sim \mathscr{N}(0, 1)$ and

Fig. 10: Homoskedastic Gaussian distributions $\mathcal{N}\left(\mu, 5^2\right)$ (drawn in blue); heteroskedastic Gaussian distributions with affine variance (drawn in green) from the family of distributions $\mathcal{N}\left(\mu, \mu\right)$; heteroskedastic multiplicative Gaussian distributions (drawn in red) from the family of distributions $\mathcal{N}\left(\mu, (0.5\mu)^2\right)$. We show the distributions with means $\mu = [2, 10, 20, 30, 40]$.

$$\sigma\left(y(x)\right) = \sqrt{ay(x) + b}, \tag{27}$$

$\sigma : \mathbb{R} \to [0, 1)$ being a univariate function (so-called *standard deviation function*[1]) that gives the signal-dependent standard deviation of the noise as a function of the deterministic noise-free signal $y(x)$. Hence, $z(x) \sim \mathcal{N}\left(y(x), \sigma^2\left(y(x)\right)\right)$, i.e.,

$$\text{pdf}\left[z(x) \mid y(x)\right](\zeta) = \frac{1}{\sigma(y(x))\sqrt{2\pi}} e^{-\left(\frac{\zeta - y(x)}{\sigma(y(x))}\right)^2}, \tag{28}$$

which constitutes a one-parameter family of distributions that depends only on the location parameter $\theta = y(x)$ that consequently defines the standard deviation of the noise. Trivially,

$$\mathrm{E}\left\{z(x \mid y(x))\right\} = y(x) \qquad \text{and} \qquad \mathrm{std}\left\{z(x) \mid y(x)\right\} = \sigma\left(y(x)\right). \tag{29}$$

It is evident that this model is extremely general and not limited to $\sigma$ in the affine-variance form (27), but can adopt arbitrary nonnegative standard deviation

---

[1] Throughout the chapter, we use the expressions *standard deviation function* and *standard deviation curve* interchangeably; similarly for the variance we consider the equivalent concepts of *variance function* or *variance curve*.
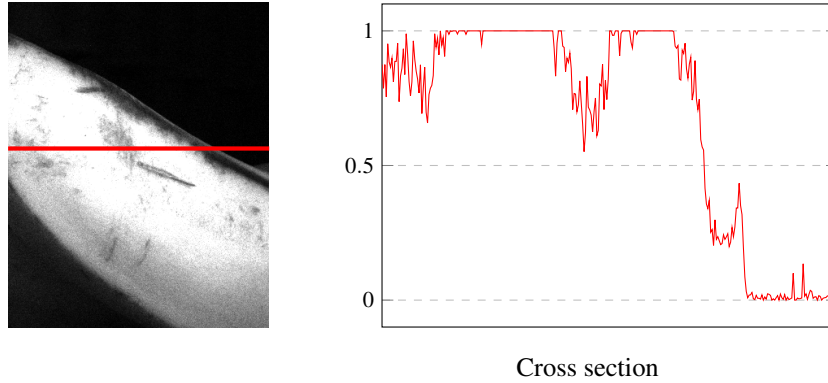
Cross section

Fig. 11: Example of a clipped noisy signal. Left: clipped detail from one of the images of the sample dataset with highlighted cross section in red. Right: plot of the cross section. Note how the signal is clipped about the boundaries $[0,1]$.

functions. We call these noise models *signal-dependent heteroskedastic Gaussian*, meaning that the variance of the Gaussian noise is not constant and depends directly on the noise-free signal. When $\sigma$ is fixed and constant, i.e., such as when $a = 0$ in (27), (26) trivially reduces to the AWGN model (2), where the noise is *signal-independent* and *homoskedastic* (constant variance).

Figure 10 illustrates three different families of distributions of the form (28): homoskedastic Gaussian distribution with constant $\sigma \equiv 5$, i.e., $\mathcal{N}\left(\mu, 5^2\right)$; heteroskedastic Gaussian distributions approximating the Poisson family, i.e., $\mathcal{N}\left(\mu, \mu\right)$, where the variance is equal to the mean; and the multiplicative noise with $\mathcal{N}\left(\mu, c\mu^2\right)$, where the standard deviation is proportional to the mean.

To clarify why the third case is multiplicative, we note that if $\sigma\left(y(x)\right) = \sqrt{c}y(x)$ we can rewrite (26) as

$$z(x) = y(x)\eta(x), \qquad \eta(x) \sim \mathcal{N}(1,c).\tag{30}$$

## 4.6 Doubly Censored Heteroskedastic Gaussian Noise: A Model for Clipped Noisy Data

All acquisition devices have a finite dynamic range that may not represent the large variation in luminosity in the scene. The device (typically at the analog-to-digital conversion stage) replaces values of intensities that exceed the range with the boundary values; in other words, the captured image can be modeled as

$$\tilde{z} = \max\left(0, \min\left(1, z\right)\right), \tag{31}$$

where the range of the captured image $\tilde{z}$ is normalized to $[0, 1]$ and where the image before the min and max operations is denoted by $z$ and is modeled as in (26). This procedure is commonly known as *clipping*. In Fig. 11 we show an example of clipped noisy signal: see how the values that exceed 0 and 1 are replaced by these bounds. The assumption that the image range in normalized to $[0, 1]$ is made, without loss of generality, for the sake of mathematical simplicity. It is clear that the noise statistics of a clipped image are not preserved by the clipping operator; in other words, if for example the acquired image $z$ is affected by Poisson-Gaussian noise, the statistics of the noise affecting the clipped image $\tilde{z}$ are not the same. Thus, $\tilde{z}$ follows a different one-parameter family of distribution and is subject to a different noise model than $z$. In what follows, we use the tilde decoration to denote variables directly related to clipped observations, following the main development and notation from Foi et al. [2008] and Foi [2009].

The corresponding noise model for the clipped observations (31) is

$$\tilde{z}(x) = \tilde{y}(x) + \tilde{\sigma}\left(\tilde{y}(x)\right)\tilde{\xi}(x), \tag{32}$$

where $\tilde{y}(x) = \mathrm{E}\{\tilde{z}(x)\}$, $\tilde{\sigma} : \tilde{y} \to \mathbb{R}^+$ gives the standard deviation of the clipped noisy data as a function of their expectation, i.e., $\tilde{\sigma}\left(\tilde{y}(x)\right) = \mathrm{std}\{\tilde{z}(x)\}$, and $\mathrm{E}\{\tilde{\xi}(x)\} = 0$, $\mathrm{std}\{\tilde{\xi}(x)\} = 1$. Because of clipping, in general, we have that

$$\tilde{y}(x) = \mathrm{E}\{\tilde{z}(x) \mid y(x)\} \neq \mathrm{E}\{z(x) \mid y(x)\} = y(x), \tag{33}$$
$$\tilde{\sigma}\left(\tilde{y}(x)\right) = \mathrm{std}\{\tilde{z}(x) \mid y(x)\} \neq \mathrm{std}\{z(x) \mid y(x)\} = \sigma\left(y(x)\right), \tag{34}$$

and $y \neq \tilde{y} \subseteq [0, 1]$. Rewriting (32) as

$$\tilde{z}(x) = y(x) + \left[\tilde{y}(x) - y(x) + \tilde{\sigma}\left(\tilde{y}(x)\right)\tilde{\xi}(x)\right],$$

we can see that, with respect to the underlying noise-free signal $y$, the clipped observations $\tilde{z}$ are corrupted by a random error (the term in square brackets) which has nonzero mean. Observe also that, even though $\mathrm{std}\{\tilde{\xi}(x)\} = \mathrm{std}\{\xi(x)\} = 1$, the distributions of $\xi$ and $\tilde{\xi}$ are different. In particular, assuming $\xi(x) \sim \mathcal{N}(0, 1)$, we have that $\tilde{\xi}(x)$ follows a doubly censored Gaussian distribution [Cohen, 1991] supported on $\left[\frac{-\tilde{y}}{\tilde{\sigma}(\tilde{y})}, \frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})}\right]$.

Figure 13 shows an example of the curves $(y, \sigma(y))$ and $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$, for $\sigma(y) = \sqrt{0.01y + 0.04^2}$. We emphasize that each curve is drawn in the corresponding expectation/standard deviation Cartesian plane (i.e., we plot the "non-clipped" $\sigma(y)$ against the $y, \sigma$ axes and the "clipped" $\tilde{\sigma}(\tilde{y})$ against the $\tilde{y}, \tilde{\sigma}$ axes). The figure illustrates the correspondence between points on the two curves given by Eqs. (31)-(33). Note that the curves from Fig. 13 are extremely similar to the scatterplot in Fig. 5, hinting that our dataset is effected by clipping.
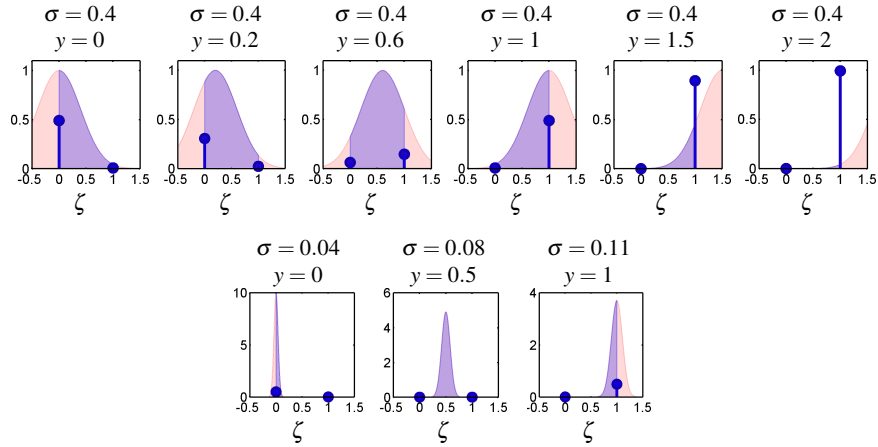
Fig. 12: Examples of doubly censored Gaussian distributions drawn in blue, and underlying uncensored Gaussian probability density function (PDF) drawn in red. Top row: the standard deviation $\sigma$ of the uncensored Gaussian PDF is fixed and equal to 0.4. Bottom row: the standard deviation of the uncensored Gaussian PDF varies according to the function $\sigma(y) = \sqrt{0.01y + 0.04^2}$, as illustrated in Fig. 13. Compare with the empirical histograms in Fig. 6.
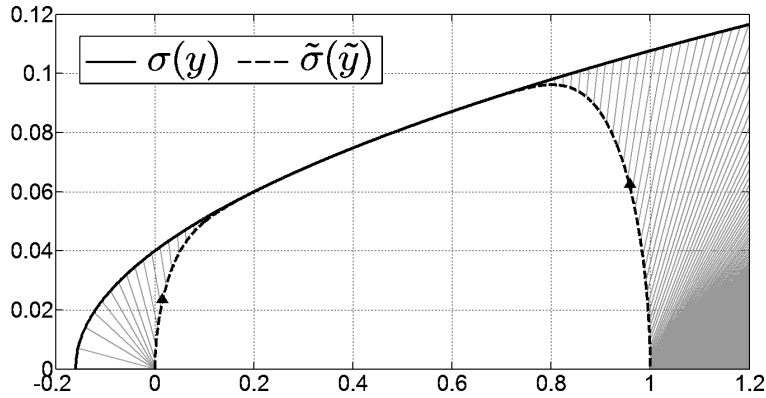


Fig. 13: Standard deviation function $\sigma(y) = \sqrt{0.01y + 0.04^2}$ (solid line) and the corresponding standard deviation curve $\tilde{\sigma}(\tilde{y})$ (dashed line). The gray segments illustrate the mapping $\sigma(y) \longmapsto \tilde{\sigma}(\tilde{y})$. The small black triangles ▲ indicate points $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$ which correspond to $y = 0$ and $y = 1$. Distributions corresponding to these curves are shown in the bottom row of Fig. 12.

### 4.6.1 Expectations and Standard Deviations of Clipped Variables and Their Transformations

A crucial point when working with clipped noisy signals is to understand how the variables and functions of the observation model (26) relate to those of the clipped observations' model (32). In particular, it is important to compute the functions $\tilde{y}$ and $\tilde{\sigma}$ given $\sigma$ and $y$, and vice versa. The PDF of the unobserved non-clipped noisy data $z \sim \mathcal{N}\left(y, \sigma^2\left(y\right)\right)$ is simply $\frac{1}{\sigma(y)}\phi\left(\frac{\zeta-y}{\sigma(y)}\right)$ (28), whereas the clipped $\tilde{z} = \max\left\{0, \min\left\{z, 1\right\}\right\}$ is distributed according to a doubly censored Gaussian distribution having a generalized PDF of the form

$$\text{pdf}\left[\tilde{z} = \zeta \mid y\right] = \Phi\left(\frac{-y}{\sigma(y)}\right)\delta_0\left(\zeta\right) + \frac{1}{\sigma(y)}\phi\left(\frac{\zeta-y}{\sigma(y)}\right)\chi_{[0,1]} + \\ \Phi\left(\frac{y-1}{\sigma(y)}\right)\delta_0\left(1-\zeta\right), \tag{35}$$

where $\chi_{[0,1]}$ denotes the characteristic function of the interval $[0,1]$ and $\delta_0$ is the Dirac delta impulse at 0. Here, $\phi$ and $\Phi$ are, respectively, the PDF and cumulative distribution function (CDF) of the standard Gaussian $\mathcal{N}\left(0,1\right)$. The first and last addends in (35) correspond to the probabilities of clipping from below and from above (under- or over-exposure), and are expressed as Dirac deltas with masses equal to the areas of the underlying Gaussian distribution that fall outside the boundaries. In Fig. 12 we give some examples of doubly censored Gaussian distributions (drawn in blue) against their uncensored counterparts, i.e., Gaussian PDFs (drawn in red): note how the impulses at the boundaries 0 and 1 have mass (shown as height) that depend on the corresponding censored parts of the Gaussian PDF outside the allowed intensity range. Thus, (35) defines a one-parameter family of distributions with parameter $y$, that also identifies the heteroskedastic Gaussian family of distributions used to build the doubly censored family.

Tedious calculations provide the following exact expressions of the expectation and variance of $\tilde{z}$ (see, e.g., [Greene, 2000], Chap. 20 or [Johnson et al., 1994]):

$$\mathrm{E}\left\{\tilde{z} \mid y\right\} = \tilde{y} = \Phi\left(\frac{y}{\sigma(y)}\right)y - \Phi\left(\frac{y-1}{\sigma(y)}\right)\left(y-1\right) + \\ \sigma\left(y\right)\phi\left(\frac{y}{\sigma(y)}\right) - \sigma\left(y\right)\phi\left(\frac{y-1}{\sigma(y)}\right), \tag{36}$$

$$\mathrm{var}\left\{\tilde{z} \mid y\right\} = \tilde{\sigma}^2\left(\tilde{y}\right) = \Phi\left(\frac{y}{\sigma(y)}\right)\left(y^2 - 2\tilde{y}y + \sigma^2\left(y\right)\right) + \tilde{y}^2 - \\ \Phi\left(\frac{y-1}{\sigma(y)}\right)\left(y^2 - 2\tilde{y}y + 2\tilde{y} + \sigma^2\left(y\right) - 1\right) + \sigma\left(y\right) \tag{37} \\ \phi\left(\frac{y-1}{\sigma(y)}\right)\left(2\tilde{y} - y - 1\right) - \sigma\left(y\right)\phi\left(\frac{y}{\sigma(y)}\right)\left(2\tilde{y} - y\right).$$

For a given function $\sigma$, these expressions explicitly define the two functions

$$\mathscr{A}_\sigma(y) = \tilde{y}, \tag{38}$$

$$\mathscr{B}_\sigma(y) = \tilde{\sigma}(\tilde{y}). \tag{39}$$

and a mapping

$$(y, \sigma(y)) \longmapsto (\tilde{y}, \tilde{\sigma}(\tilde{y})) \tag{40}$$

that brings the standard deviation curve $(y, \sigma(y))$ to its clipped counterpart $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$. In Fig. 13, we give an example of standard deviation function $\sigma(y) = \sqrt{0.01y + 0.04^2}$ (solid line) and clipped standard deviation curve $\tilde{\sigma}(\tilde{y})$ (dashed line). The gray segments are used to illustrate the mapping $\sigma(y) \mapsto \tilde{\sigma}(\tilde{y})$ (40).

The inverse of $\mathscr{A}_\sigma$ will be formally denoted as

$$\mathscr{C}_\sigma : \tilde{y} \longmapsto y = \mathscr{C}_\sigma(\tilde{y}). \tag{41}$$

Invertibility requires some hypotheses on the standard deviation function $\sigma$; for instance, it can be shown that (41) is well defined provided that $\sigma(y) = \sqrt{ay + b}$ with $a > 0$ and $-\frac{b}{a} \leq \frac{1}{2}$ [Foi, 2009]. The function $\mathscr{B}_\sigma$ is instead not invertible for the most common types of standard deviation functions, for which $\tilde{\sigma}(0) = \tilde{\sigma}(1) = 0$ [Foi, 2009].

Although the expressions (36) and (37) can be eventually useful for a numerical implementation, they are cumbersome and cannot be easily manipulated for further analysis.

### 4.6.2 Approximation Using Singly Censored Variables

We can simplify the above analysis under the assumption that there are no values of $y$ for which the clipping from below ($z < 0$, $\tilde{z} = 0$) and clipping from above ($z > 1$, $\tilde{z} = 1$) may both occur with significant probabilities. This assumption means that, for a fixed $y$, at most one of the impulses in the PDF (35) has mass appreciably larger than 0. In practice, this assumption is always verified under normal capture settings. Exceptions are extreme situations where the noise is dramatically strong with respect to the $[0,1]$ range (e.g., $\sigma(y) \gg 0.2$ for all $y \in [0,1]$, like in the case illustrated in Fig. 12).

Let $v \sim \mathcal{N}(\mu, 1)$ be a normally distributed random variable with mean $E\{v\} = \mu$ and unitary variance, and let $\tilde{v} = \max\{0, v\}$. Similar to (36) and (37), it can be shown that the expectation $E\{\tilde{v}\}$ and the variance $\text{var}\{\tilde{v}\}$ of the clipped (from below) $\tilde{v}$ are

$$E\{\tilde{v}\} = \mathscr{E}_m(\mu) = \Phi(\mu)\mu + \phi(\mu), \tag{42}$$

$$\begin{aligned} \text{var}\{\tilde{v}\} = \mathscr{S}_m^2(\mu) &= \Phi(\mu) + \mathscr{E}_m(\mu)\mu - \mathscr{E}_m^2(\mu) = \\ &= \Phi(\mu) + \phi(\mu)\mu - \phi^2(\mu) + \\ &+ \Phi(\mu)\mu(\mu - \Phi(\mu)\mu - 2\phi(\mu)). \end{aligned} \tag{43}$$
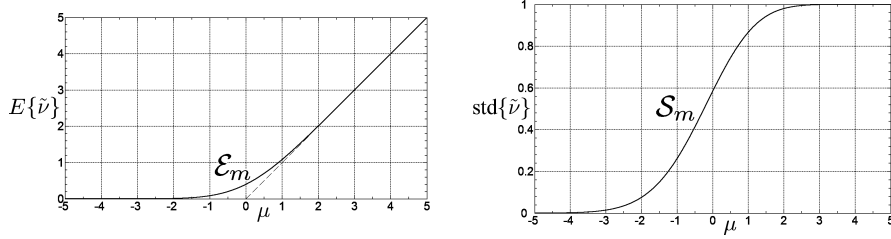
Fig. 14: Expectation $E\{\tilde{v}\}$ and standard deviation $\text{std}\{\tilde{v}\}$ of the clipped $\tilde{v} = \max\{0, v\}$ as functions $\mathscr{E}_m$ and $\mathscr{S}_m$ of $\mu$, where $\mu = E\{v\}$ and $v \sim \mathscr{N}(\mu, 1)$.
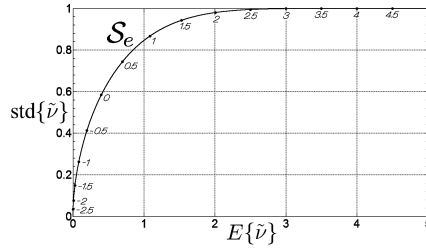


Fig. 15: Standard deviation $\text{std}\{\tilde{v}\}$ of the clipped $\tilde{v} = \max\{0, v\}$ as function $\mathscr{S}_e$ of the expectation $E\{\tilde{v}\}$. The numbers in italic indicate the corresponding value of $\mu$, where $\mu = E\{v\}$ and $v \sim \mathscr{N}(\mu, 1)$.

The plots of the expectation $E\{\tilde{v}\} = \mathscr{E}_m(\mu)$ and of the standard deviation $\text{std}\{\tilde{v}\} = \mathscr{S}_m(\mu)$ are shown, as functions of $\mu$, in Fig. 14. Figure 15 combines these two functions and visualizes the mean-standard deviation curve characteristic of standardized clipped variables.

Exploiting these functions, the direct and inverse transformations which link $\sigma$ and $y$ to $\tilde{y}$ and $\tilde{\sigma}$ can be expressed in the following compact forms [Foi et al., 2008].

*Direct transformation: obtain $\tilde{y}$ and $\tilde{\sigma}$ from $y$ and $\sigma$*

Provided that $y = E\{z\}$ and $\sigma(y) = \text{std}\{z\}$ from the basic model (32) are known, the expectation $\tilde{y} = E\{\tilde{z}\}$ and the standard deviation $\tilde{\sigma}(\tilde{y}) = \text{std}\{\tilde{z}\}$ from the observation model (32) are obtained as

$$\tilde{y} = \mathscr{A}_\sigma(y) \approx \mathscr{A}(y, \sigma(y)) =$$
$$= \sigma(y)\mathscr{E}_m\left(\tfrac{y}{\sigma(y)}\right) + 1 - y - \sigma(y)\mathscr{E}_m\left(\tfrac{1-y}{\sigma(y)}\right), \tag{44}$$
$$\tilde{\sigma}(\tilde{y}) = \mathscr{B}_\sigma(y) \approx \mathscr{B}(y, \sigma(y)) = \sigma(y)\mathscr{S}_m\left(\tfrac{y}{\sigma(y)}\right)\mathscr{S}_m\left(\tfrac{1-y}{\sigma(y)}\right). \tag{45}$$

Compared to the exact (38) and (39), the approximate equations (44) and (45) provide a more intuitive description of the transformations that bring the standard deviation curve $(y, \sigma(y))$ to its clipped counterpart $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$. For instance, provided $y$ is
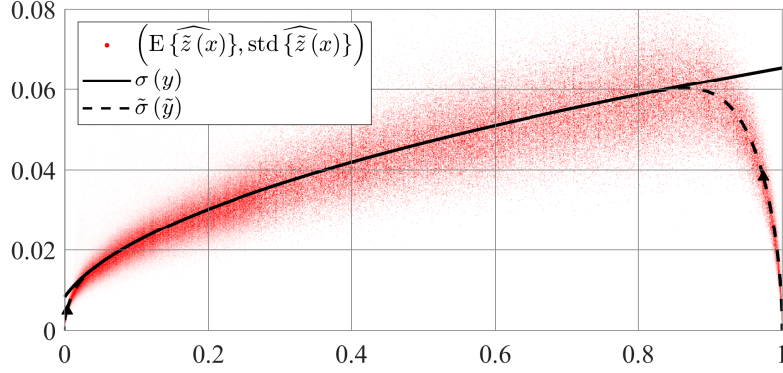
Fig. 16: Scatterplot of the pairs $\left(\mathrm{E}\{\widehat{\tilde{z}(x)}\}, \mathrm{std}\{\widehat{\tilde{z}(x)}\}\right)$ drawn as red dots, and estimated clipped (black dashed line) and non-clipped (black continuous line) noise standard deviation curves. The small black triangles ▲ indicate points $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$ which correspond to $y = 0$ and $y = 1$.

sufficiently smaller than 1, by observing Fig. 14 it is easy to realize that $\mathscr{E}_m\left(\frac{1-y}{\sigma(y)}\right)$ and $\mathscr{S}_m\left(\frac{1-y}{\sigma(y)}\right)$ can be substituted by $\frac{1-y}{\sigma(y)}$ and 1, respectively (the substitution is asymptotically exact). Thus, for describing the clipping from below, (44) and (45) can be reduced to, respectively, $\sigma(y)\mathscr{E}_m\left(\frac{y}{\sigma(y)}\right)$ and $\sigma(y)\mathscr{S}_m\left(\frac{y}{\sigma(y)}\right)$, which allows to construct the graph of $(\tilde{y}, \tilde{\sigma}(\tilde{y}))$ in the vicinity of $(0, 0)$ by simple manipulations of the graphs of $\mathscr{E}_m$ and $\mathscr{S}_m$.

*Inverse transformation: obtain y from $\tilde{\sigma}$ and $\tilde{y}$*

The approximation of (41) for calculating the non-clipped $y$ (26) from the clipped $\tilde{y}$ and $\tilde{\sigma}(\tilde{y})$ can be given as

$$y = \mathscr{C}_\sigma(\tilde{y}) \approx \mathscr{C}(\tilde{y}, \tilde{\sigma}(\tilde{y})) =$$
$$= \tilde{y}\mathscr{E}_r\left(\frac{\tilde{y}}{\tilde{\sigma}(\tilde{y})}\right) - \tilde{y} + 1 - (1 - \tilde{y})\mathscr{E}_r\left(\frac{1-\tilde{y}}{\tilde{\sigma}(\tilde{y})}\right), \tag{46}$$

where $\mathscr{E}_r$ is defined implicitly as function of $\rho = \frac{\mathscr{E}_m(\mu)}{\mathscr{S}_m(\mu)} = \frac{\mathrm{E}\{\tilde{v}\}}{\mathrm{std}\{\tilde{v}\}}$ by $\mathscr{E}_r(\rho) = \frac{\mu}{\mathscr{E}_m(\mu)}$.

## 5 Estimation of the Standard Deviation Curve

The main purpose of noise estimation algorithms is to estimate the standard deviation curve. The most common framework first builds a scatterplot with mean values of the signal on the abscissa, and corresponding standard deviations (variances) on
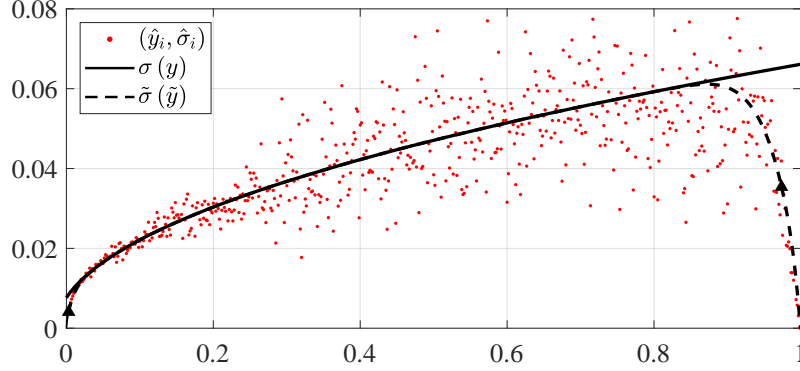
Fig. 17: Noise standard deviation (black curve) $\sigma(y)$ estimated from one image from the dataset. We show also the estimate of the clipped standard deviation (dashed curve) $\tilde{\sigma}(\hat{y})$ and the scatterplot used for the fitting. For the estimation we used the algorithm [Azzari and Foi, 2014a], and the estimated parameters are $a = 4.315 \cdot 10^{-3}$ and $b = 5.814 \cdot 10^{-5}$.

the ordinate (see the scatterplot in Fig. 5); then, it fits a parametric curve over these points. When only a single image is used for estimation, it is common practice to compute each scatterpoint from homogeneous samples, i.e., each element in a sample shares a unique common expectation value (hence, they also share a common unique variance value). This practice is based on the fact that the sample variance of homogeneous samples is an unbiased estimator of the noise variance for that particular expectation value. Consequently, each point in the scatterplot has a direct relation to a point on the curve we want to estimate.

Regardless of the estimation algorithm, it is convenient to model each mean-standard deviation pair estimate $(\hat{y}_i, \hat{\sigma}_i)$ with a bivariate PDF of the form

$$\text{pdf}\left[(\hat{y}_i, \hat{\sigma}_i) \,|\, \tilde{y}_i = \tilde{y}\right] = \text{pdf}\left[\hat{y}_i | \tilde{y}_i = \tilde{y}\right] \text{pdf}\left[\hat{\sigma}_i | \tilde{y}_i = \tilde{y}\right], \qquad (47)$$

where the subscript $i$ indicates the generic $i$-th estimated pair in the scatterplot. Note that the joint probability is reduced to a product of univariate distributions because one can assume that the estimates $\hat{y}_i$ and $\hat{\sigma}_i$ are independent [Johnson et al., 1994]. Given the distributions of all the pairs $\{\hat{y}_i, \hat{\sigma}_i\}_{i=1}^N$, the posterior likelihood function $L$ can be calculated as the product of all the densities $\text{pdf}\left[(\hat{y}_i, \hat{\sigma}_i) \,|\, \tilde{y}_i = \tilde{y}\right]$ with the prior density $\text{pdf}[y]$ of $y$:

$$L(\theta) = \prod_{i=1}^{N} \int \text{pdf}\left[(\hat{y}_i, \hat{\sigma}_i) \,|\, \tilde{y}_i = \tilde{y}\right] \text{pdf}[y] \, dy, \qquad (48)$$

where $\theta$ is an $m$-dimensional vector composed by the model parameters to be estimated. The likelihood function $L$ expresses the joint probability function of the whole collection $\{\hat{y}_i, \hat{\sigma}_i\}_{i=1}^N$. The vector $\theta$ determines univocally the standard devi-

ation curve $\sigma(y)$, e.g., $\theta = [a, b]$, $m = 2$ in case of the affine mean-variance relation (27). The maximization of $L$ leads to the estimation of the noise curve parameters $\theta$. Note that although clipped data is bound to the $[0, 1]$ interval, assuming that pdf $[y]$ is a uniform density supported over $[0, 1]$ is incorrect, since this prior probability refers to the signal before clipping and this signal can naturally exceed the boundaries of the unit interval, e.g., an overexposed signal $y$ is typically larger than 1. Therefore, one may assume either a noninformative prior for which all $y \in \mathbb{R}$ are equiprobable or most commonly a nonnegative signal prior for which all $y \geq 0$ are equiprobable and $y < 0$ is impossible, which simplifies (48) to

$$L(\theta) = \prod_{i=1}^{N} \int_{0}^{\infty} \text{pdf}\left[(\hat{y}_i, \hat{\sigma}_i) \,|\, \tilde{y}_i = \tilde{y}\right] dy. \tag{49}$$

The joint PDF can be, for example, a product of two univariate Gaussian PDFs (see, e.g., [Foi et al., 2008]), a product of Gaussian-Cauchy mixtures [Azzari and Foi, 2014a], or other PDFs depending on the data and on the method adopted.

Figure 16 shows an example of scatterplot fitting through maximization of the likelihood function (49): the scatterpoints, drawn as red dots, are the pairs $\left(\widehat{\text{E}\{\tilde{z}(x)\}}, \widehat{\text{std}\{\tilde{z}(x)\}}\right)$ estimated in (8) and (9) using our dataset (see the same scatterplot in Fig. 5); the estimated lines are drawn in black and have been estimated maximizing (49) using the PDFs introduced in (10)-(11), where the index $i$ takes the place of the index $x$. Note how the estimate of the clipped standard deviation (dashed line) diverges from the non-clipped one (continuous line) when approaching the boundaries $[0, 1]$, which follows the divergence between the basic heteroskedastic Gaussian distribution and the proper heteroskedastic singly censored Gaussian (e.g., as illustrated in Fig. 12).

Most algorithms in the literature for estimating the variance function $\sigma^2$ are designed for an affine function of the signal mean, because it provides a reasonably accurate description of the output (without clipping) of imaging sensors commonly found in digital cameras. Nonetheless, some algorithms such as [Sutour et al., 2015; Azzari and Foi, 2014a], inherently support models of arbitrary order (e.g., a variance function defined as a quadratic or higher-order polynomial of the signal mean) and can be used in scenarios where a higher-order of approximation of the noise variance function is needed, such as the case considered in Sect. 7.

In the remainder of this section, we describe relevant approaches for the estimation of the noise standard deviation curve (or equivalently the variance curve) under the assumption of white noise. While the problem of estimating the correlation (power spectral density) of colored noise is investigated in Sect. 6.3.2, the majority of the methods below can be extended, with due modifications, to the more generic case of the estimation of the standard deviation curve of signal-dependent colored noise. Note that our overview is without any pretension of completeness, with the only goal of briefly introducing the fundamentals behind the most popular approaches for noise estimation.

## 5.1 Patch-Based Methods

A prominent *patch-based* algorithm is introduced by Lee and Hoppel [1989]. The algorithm divides the image into small patches and computes their mean and variance to build the scatterplot of mean-variance pairs. Since image blocks might contain heterogeneous elements that would mislead the estimation of the local variances, the authors estimate the noise parameters by finding the curve that intersects most of the scatterpoints. In this way they reduce the effect of outliers that usually appear far from the majority of the scatter-points. In a similar work, Amer and Dubois [2005] evaluate, using directional derivative filters, the uniformity of each patch that generated a data point. Comparing the uniformity against a threshold, the algorithm decides whether to use the scatterpoint (if the patch elements are homogeneous) or to discard it. Finally, since the outliers have been already excluded, a simple least square (LS) fitting is adopted. In [Sutour et al., 2015], the authors divide the image in nonoverlapping blocks; based on the Kendall's $\tau$ coefficients, adopted to find the correlation between elements from the same block, the blocks are then classified as homogeneous or heterogeneous. The heterogeneous blocks are discarded, while the homogeneous ones are used to compute the local statistics for the fitting of the noise variance curve. An important aspect of the algorithm is that a robust fitting is performed by minimizing the $\ell_1$ error of the residuals. Similar works can be found in [Meer et al., 1990; Lee, 1981; Mastin, 1985]; however, we decide not to go into further details.

An interesting variant has been proposed by Boulanger et al. [2010]. They divide the image into adaptive blocks whose size depends on the variance of their elements (homogeneity). If the variance of a block matches the variance model (Fisher test is used to compare the two), then the block is considered homogeneous, otherwise the block is further split into four parts and each *sub*block is then analyzed as before. Finally, the authors perform noise parameters estimation via robust linear regression of the local estimates.

## 5.2 Segmentation-Based Methods

We now describe the most relevant *segmentation-based* approaches for noise estimation. Gravel et al. [2004] segment the observed noisy image into homogeneous samples that are each used to compute a scatter-point. The segmentation is performed by first smoothing the observed image, and then by grouping pixels with similar intensity. This leverages the fact that a smoothing operator suppresses the noise and facilitates the segmentation process. Pixels from edges and texture are excluded from the estimation, since the segmentation is inaccurate in those regions. The noise parameters are finally estimated using a weighted regression of the scatterplot points.

Another type of segmentation is proposed in [Liu et al., 2006], where the authors do not filter the image, but they bin the image elements using a K-means clustering

method. A robust fitting algorithm is then adopted to cope with possible inaccuracies of the K-means clustering: the noise parameters are estimated by fitting a lower envelope of the scatterplot, computed by maximizing a likelihood function (see (49)) that takes into consideration the possible overestimation of the scatterpoints variances. Similarly, Foi et al. [2008] filter the observed image, segment it, and then maximize a likelihood function to estimate the noise parameters. A major novelty introduced by this work is that it takes under consideration the clipping of the data. Figure 17 shows the estimation results of the algorithm [Foi et al., 2008] applied to an image of our dataset. Note how, although the number of scatterpoints used for the likelihood maximization is much smaller compared to the scatterpoints in Fig. 16, the estimated curves are quite accurate. The algorithm deals with clipped data using the mappings (45), (44), and (46), and estimates a parametric model of the noise affecting the non-clipped signal from the clipped noisy observation, simultaneously providing the standard deviation functions $\tilde{\sigma}$ for the clipped noise and $\sigma$ for the underlying data before clipping.

## 5.3 Alternative Approaches

In [Azzari and Foi, 2014b], it is shown that the use of homogeneous samples is not a requirement to estimate the affine-variance model (27), and thus the noise parameters may be estimated without any image segmentation and by leveraging instead robust filters.

The algorithm described in [Mäkitalo and Foi, 2014] estimates the noise by exploiting the variance stabilization achieved by the generalized Anscombe transformation (GAT). The GAT is a nonlinear univariate transformation that converts signal-dependent Poisson-Gaussian noise into approximately additive and signal-independent noise with unitary variance. The GAT is defined based on the same parameters $a$ and $b$ adopted to describe the Poisson-Gaussian noise. This approach for noise estimation is based on the observation that the variance stabilization is not accurate when the GAT parameters do not match the ones from the noise model. Hence, the noise parameters are obtained by finding the GAT parameters that yield the most accurate variance stabilization.

Finally, the algorithm in [Pyatykh et al., 2013] estimates the noise parameters by analyzing the last eigenvalues of the singular value decomposition (SVD) of homogeneous patches from a noisy image. These eigenvalues capture noise and virtually no signal, and are therefore used to estimate accurately the noise parameters.

## 6 Correlated Noise

So far we have described noise models characterized by a flat power spectral density, i.e., various types of white noise. According to those models, the noise affecting dif-

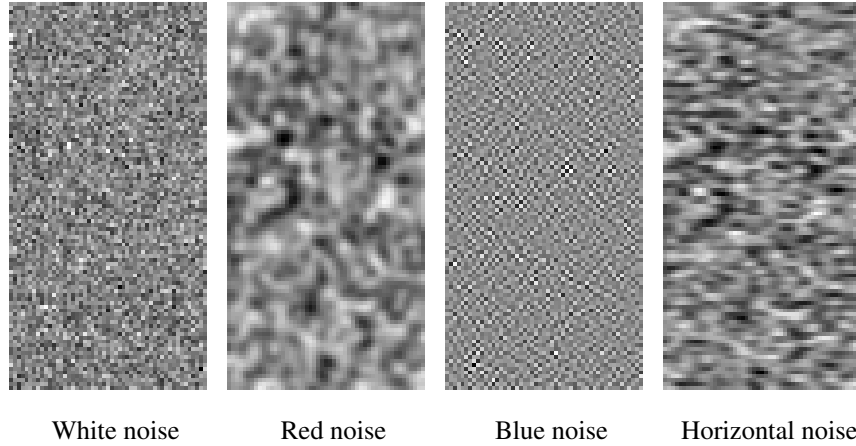White noise          Red noise          Blue noise          Horizontal noise

Fig. 18: White noise vs. three examples of colored noise.

ferent pixels is uncorrelated. However, in many practical applications, there could be correlation in the errors: this correlation might be due to the physics of the acquisition system, to the sensor's readout process, to cross-talk between neighboring pixels, or due to processing performed on the raw image after the acquisition. These types of acquisition errors can be modeled by the so-called *colored* noise models, which assume a stationary spatial correlation among noise realizations, as illustrated by the examples in Fig. 18, and that can be characterized by a nonconstant power spectral density of the noise. In this section, we discuss models for correlated noise, starting from the simpler case of signal-independent stationary colored noise and then introducing two forms of signal-dependent colored noise.

## 6.1 Stationary Correlated Noise

The generic model for a noisy image $z$ corrupted by a zero-mean stationary additive correlated noise $\eta$ is

$$z(x) = y(x) + \eta(x), \qquad \eta(x) = (\nu \circledast g)(x), \tag{50}$$

where $x \in \Omega \subset \mathbb{Z}^2$, $y$ is a deterministic noise-free image, $\nu$ is zero-mean independent noise with unit variance, and $g$ is a convolution kernel that determines the variance and the spatial correlation of the noise $\eta$. Specifically, the variance and the power spectrum (PSD) (4) of $\eta$ are

$$\text{var}\{\eta(x)\} = \|g\|_2^2, \qquad \Psi = \text{var}\{\mathscr{F}[\eta]\} = |\mathscr{F}[g]|^2 |\Omega|, \tag{51}$$

and by Parseval's isometry we have

$$\text{var}\{\eta(x)\} = |\Omega|^{-1} \|\mathscr{F}[g]\|_2^2 = |\Omega|^{-2} \|\Psi\|_1. \tag{52}$$

For the sake of simplicity, we assume a circulant convolution in order to leverage the convolution theorem. Model (50) reduces to the AWGN model (2) when $g$ is a scaled Dirac impulse with mass $\sigma$ and $v$ is Gaussian.

We say that the noise is *colored* when the noise power spectrum is markedly not flat. Whenever there is a dominant spectral band characterized by a significantly larger noise variance, the spectral position of this band determines the "color" of the noise; thus, noise predominately affecting the low frequencies is often called "red noise", as opposed to a "blue noise" which is mostly localized on the high frequencies. Figure 18 compares white noise with three examples of correlated noise. Different types of spatial correlation can be appreciated, with the noise affecting a pixel influenced by the surrounding noise realizations.

## 6.2 Correlated Signal-Dependent Noise Model

Depending on the physics and hardware of the acquisition process, noise can feature both correlation and signal-dependent characteristics. However, a PSD and a variance function cannot be defined exactly within the same model, because their underlying generative processes are incompatible with each other. The two models presented in the next subsections show two extremes of a compromise: the first model ignores the means of neighboring pixels and thus the variance is defined exactly at the expense of the PSD model, which is only approximate; in the second model the PSD is expressed exactly while the variance function is approximate. In intermediate cases, neither the PSD nor the variance function may be defined exactly. Although formally quite different, the model discrepancies are shown to be typically small in smooth areas of the image, making these approximate models useful and constructive [e.g., Azzari and Foi, 2017; Borges et al., 2017].

### 6.2.1 Noise Scaling Post Correlation

The first model assumes that the signal-dependent part of the noise acts as a deterministic scaling term to a stationary correlated noise:

$$\begin{aligned} z(x) &= y(x) + \sigma(y(x))\,\eta(x), \\ \eta &= v \circledast g, \qquad v(\cdot) \sim \mathscr{N}(0,1), \qquad \sigma : y \to \mathbb{R}^+, \end{aligned} \tag{53}$$

where $\sigma$ is a generic standard deviation function. The expectation and variance of $z$ are, respectively,

$$\mathrm{E}\{z\} = y, \tag{54}$$

$$\mathrm{var}\{z\} = \mathrm{var}\{\sigma(y)\,v \circledast g\} = \sigma^2(y)\,\mathrm{var}\{v \circledast g\} =$$
$$\sigma^2(y)\,\|g\|_2^2 = \sigma^2(\mathrm{E}\{z\})\,\|g\|_2^2. \tag{55}$$

With regard to the PSD, it can be approximated as

$$\mathrm{var}\{\mathscr{F}[z]\} \approx |\mathscr{F}[g]|^2\,\left\|\sigma^2(y)\right\|_1. \tag{56}$$

Roughly speaking, (53) describes a physical process where the correlating process takes place before the signal-dependent amplification of errors.

### 6.2.2 Noise Scaling Prior to Correlation

The second model considers instead a case where the correlating process operates on a ready signal-dependent white noise model:

$$z'(x) = y(x) + \sigma(y(x))\,v(x), \tag{57}$$

$$z(x) = (z' \circledast g)(x). \tag{58}$$

The expected value and variance of $z$ are, respectively,

$$\mathrm{E}\{z\} \approx \mathrm{E}\{z'\}\,\|g\|_1 = y\,\|g\|_1, \tag{59}$$

$$\mathrm{var}\{z\} \approx \mathrm{var}\{z'\}\,\|g\|_2^2 = \sigma^2(y(x))\,\|g\|_2^2, \tag{60}$$

where the approximations become accurate in large smooth areas of the image where the intensity changes gradually. Combining (59) and (60) we obtain

$$\mathrm{var}\{z\} \approx \sigma^2\left(\frac{\mathrm{E}\{z\}}{\|g\|_1}\right)\|g\|_2^2. \tag{61}$$

The PSD for (58) is

$$\mathrm{var}\{\mathscr{F}[z]\} = |\mathscr{F}[g]|^2\,\left\|\sigma^2(y)\right\|_1 \approx |\mathscr{F}[g]|^2\,\left\|\sigma^2\left(\frac{\mathrm{E}\{z\}}{\|g\|_1}\right)\right\|_1. \tag{62}$$

Thus, both Model 1 and Model 2 express the variance of $z$ as a function of its expectation, where the main differences consist merely in a scaling of the variables, and this scaling is determined by the $\ell_1$ and $\ell_2$ norms of the convolution kernel $g$. Therefore, macroscopically, the two models are comparable.

Before we proceed, let us observe that the degree of approximation in (59)-(60) can be quantified by expanding the expression of $z(x_0)$ at the generic coordinate $x_0$:

$$z(x_0) = [z' \circledast g](x_0) = \sum_{x \in \Omega} z'(x_0 - x)\,g(x). \tag{63}$$

The expected value and variance of the generic pixel $z(x_0)$ can be thus calculated, respectively, as

$$\mathrm{E}\{z\}(x_0) = \sum_{x \in \Omega} \mathrm{E}\{z'(x_0 - x)\} g(x) = \sum_{x \in \Omega} y(x_0 - x) g(x) =$$

$$\sum_{x \in \Omega} \sum_{k=0}^{+\infty} \frac{\partial^k y(x_0) x^k}{(-1)^k k!} g(x) = \sum_{k=0}^{+\infty} \frac{\partial^k y(x_0)}{(-1)^k k!} \sum_{x \in \Omega} x^k g(x), \qquad (64)$$

$$\mathrm{var}\{z\}(x_0) = \sum_{x \in \Omega} \mathrm{var}\{z'(x_0 - x)\} g(x) = \sum_{x \in \Omega} \sigma^2(x_0 - x) g(x) =$$

$$\sum_{x \in \Omega} \sum_{k=0}^{+\infty} \frac{\partial^k (\sigma^2 \circ y)(x_0) x^k}{(-1)^k k!} (g(x))^2 =$$

$$\sum_{k=0}^{+\infty} \frac{\partial^k (\sigma^2 \circ y)(x_0)}{(-1)^k k!} \sum_{x \in \Omega} x^k (g(x))^2, \qquad (65)$$

where the final expressions of (64)-(65) come from the Maclaurin expansions of $y(x_0 - x)$ and $\sigma^2(x_0 - x)$ at the generic coordinate $x_0 \in \Omega$, and $\partial^k (\sigma^2 \circ y)(x)$ is the $k$th derivative of the composite function $\sigma^2(y(x))$, and it can be computed, for example, with the formula of Faà Di Bruno [1857].

For approximation of order zero, i.e., when $k = 0$, (64)-(65) reduce to (59)-(60). Thus, in regions where $y$ is constant, the approximations are always valid. If we assume $g$ even symmetric, both $\sum_{x \in \Omega} x^k g(x)$ and $\sum_{x \in \Omega} x^k (g(x))^2$ vanish when $k$ is odd since $x^k$ is odd symmetric. Therefore, for even symmetric kernels $g$, the approximations (59)-(60) differ from (64)-(65) only for approximation terms of $y(x_0 - \cdot)$ and $\sigma^2(x_0 - \cdot)$ of even order 2 or greater.

In case of affine noise variance (22), i.e. $\sigma^2(y(x)) = ay(x) + b$, (65) becomes

$$\mathrm{var}\{z\}(x_0) = (ay(x_0) + b) \sum_{x \in \Omega} (g(x))^2 -$$

$$a(\partial y)(x_0) \sum_{x \in \Omega} x^k (g(x))^2 +$$

$$a \sum_{k=2}^{+\infty} \frac{(\partial^k y)(x_0)}{(-1)^k k!} \sum_{x \in \Omega} x^k (g(x))^2. \qquad (66)$$

## 6.3 Estimation

### 6.3.1 Estimation of the Noise Standard Deviation Function

The methods for estimating the standard deviation curve which were briefly reviewed in Sect. 5 are formally designed for independently distributed noise, i.e., white noise. If these methods are applied to correlated noise, they may fail to estimate correctly the curve. In this regard, we note that many of these methods (e.g.,

Gravel et al. [2004]; Foi et al. [2008]; Azzari and Foi [2014b]; Pyatykh et al. [2013]; Mäkitalo and Foi [2014]) employ some kind of high-pass filtering to reduce the impact of the signal $y$ on the estimation of noise variance. Whereas white noise affects in equal manner different frequency bands, correlated noise is distributed unevenly over the frequency spectrum: without prior knowledge of the noise power spectrum, there is uncertainty about the proportion of noise that is effectively maintained after high-pass filtering, hence a potential risk of significant overestimation or underestimation of the curve. To understand this phenomenon, let us consider the observation model (53) and assume that scatterpoints are obtained from estimating the variance of the output of filtering $z$ by a high-pass kernel $h$. Restricting the estimation over a sufficiently large region where $y$ is homogeneous, we can treat $\sigma(y)$ as a constant and we have

$$\begin{aligned}
\operatorname{var}\{z \circledast h | y\} &= \operatorname{var}\{\sigma(y)\,\eta \circledast h\} = \sigma^2(y) \operatorname{var}\{v \circledast g \circledast h\} = \\
&= \sigma^2(y) \|g \circledast h\|_2^2 = \sigma^2(y)\,|\Omega|^{-1}\,\|\mathscr{F}[g \circledast h]\|_2^2 = \\
&= \sigma^2(y)\,|\Omega|^{-1}\,\|\mathscr{F}[g]\,\mathscr{F}[h]\|_2^2 = \\
&= \sigma^2(y)\,|\Omega|^{-2}\,\left\|\Psi\,|\mathscr{F}[h]|^2\right\|_1.
\end{aligned} \tag{67}$$

Thus, even when $\|h\|_2 = 1$, it may happen that $\operatorname{var}\{z \circledast h | y\} \neq \operatorname{var}\{z|y\} = \sigma^2(y)\,|\Omega|^{-2}\,\|\Psi\|_1$, with various outcomes depending on how $\mathscr{F}[h]$ correlates with the PSD $\Psi$. In particular,

$$\operatorname{var}\{z \circledast h | y\} = \operatorname{var}\{z|y\}\,\left\|\Psi\,|\mathscr{F}[h]|^2\right\|_1 \|\Psi\|_1^{-1}. \tag{68}$$

This means that if all scatterpoints are estimated using a unique filter $h$, we can estimate the noise variance function modulo a scaling factor that depends on the PSD $\Psi$. As we show next, the fact that this scaling factor is unknown does not prevent estimating the PSD $\Psi$, from which we can then resolve (68) and obtain an estimate of the proper noise variance function.

### 6.3.2 PSD Estimation

The typical approach to estimating the noise PSD consists of processing the standardized noise with a filter bank of band-pass filters, producing a collection of sub-band images and by calculating the sample variance (or a robust estimate of the variance) of the noise in these subbands. Modulo a, possibly unknown, scaling factor $\lambda$, standardized noise can be represented in the form

$$\check{\eta} = \lambda v \circledast \check{g}, \tag{69}$$

where $\check{g} = g\,\|g\|_2^{-1}$, $v$ is zero-mean independent noise with unit variance, and hence $\check{\eta}$ has zero-mean and variance $\lambda^2$.

The signal-independent colored noise model (50) already has the form (69) with $\check{\eta} = \eta$ and $\lambda = \|g\|_2$. Signal-dependent colored noise of the forms (53) can be
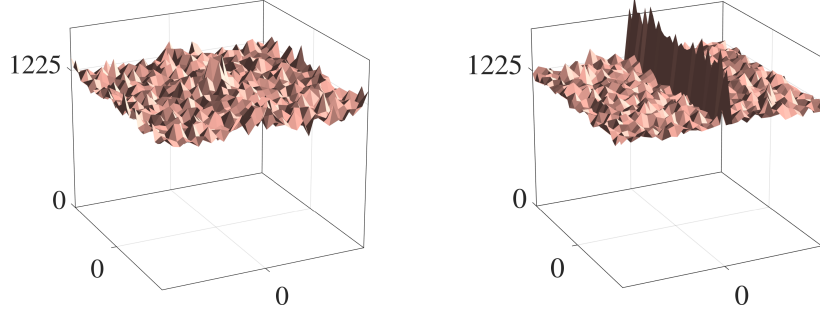
Fig. 19: 35×35 Fourier-domain PSD estimated for a Samsung S5K2L2 CMOS sensor from the dataset of Section (3) (left) and for a Foveon X3 sensor (right). The former is practically flat (thus representing white noise), whereas the latter shows a mild nonuniformity characteristic of either row or column striping noise.

reduced to (69) in several ways, e.g., by standardization of the samples with respect to the noise model fitted to the mean, or by variance stabilizing transformations [Starck et al., 1998; Mäkitalo and Foi, 2014]. If the standardization is obtained via a noise variance function estimated modulo a scaling factor as in (68), then we arrive at (69) with $\lambda^2 = \left\| \Psi \left| \mathscr{F}[h] \right|^2 \right\|_1^{-1} \|\Psi\|_1$.

Let us consider a set of band-pass filters $\mathscr{H} = \{h_1, \ldots, h_J\}$. The generic output from the $j$th filter is

$$\breve{\eta} \circledast h_j = (\lambda \nu \circledast \breve{g}) \circledast h_j,  \tag{70}$$

and its variance

$$\mathrm{var}\left\{ \breve{\eta} \circledast h_j \right\} = \lambda^2 \mathrm{var}\left\{ \nu \circledast \breve{g} \circledast h_j \right\} = \lambda^2 \mathrm{var}\left\{\nu\right\} \left\| \breve{g} \circledast h_j \right\|_2^2 = \lambda^2 \left\| \breve{g} \circledast h_j \right\|_2^2.  \tag{71}$$

By Parseval's isometry and (51), this variance takes the form

$$\mathrm{var}\left\{ \breve{\eta} \circledast h_j \right\} = \lambda^2 \left| \Omega \right|^{-2} \left\| \left| \mathscr{F}[h_j] \right|^2 \breve{\Psi} \right\|_1,  \tag{72}$$

where $|\Omega| = N_1 N_2$ denotes the cardinality of $\Omega$ and $\breve{\Psi}$ is a normalized PSD linked to the normalization of the kernel $\breve{g}$,

$$\breve{\Psi} = \Psi \|g\|_2^{-1} = |\Omega|^2 \Psi \|\Psi\|_1^{-1} \quad \text{and} \quad \left\| \breve{\Psi} \right\|_1 = |\Omega|^2.  \tag{73}$$

Hence, $\left\| \lambda^2 \breve{\Psi} \right\|_1 = \lambda^2 |\Omega|^2$, which shows that estimating $\lambda^2 \breve{\Psi}$ trivially yields also separate $\lambda = |\Omega|^{-1} \sqrt{\left\| \lambda^2 \breve{\Psi} \right\|_1}$ and consequently $\breve{\Psi}$.

Therefore, the problem at hand amounts to estimating $\lambda^2 \breve{\Psi}$ from $\mathrm{var}\left\{ \breve{\eta} \circledast h_j \right\}$, $j = 1, \ldots, J$. To this end, it is convenient to rewrite (72) in matrix notation as

$$\mathbf{V} = \mathbf{F}^T \mathbf{P}  \tag{74}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathrm{var}\{z_1\} \\ \vdots \\ \mathrm{var}\{z_J\} \end{bmatrix}, \qquad \mathbf{P} = \lambda^2 \begin{bmatrix} \check{\Psi}(1) \\ \vdots \\ \check{\Psi}(|\Omega|) \end{bmatrix} \tag{75}$$

$$\mathbf{F} = \frac{1}{|\Omega|^2} \begin{bmatrix} |\mathscr{F}[h_1](1)|^2 & \cdots & |\mathscr{F}[h_J](1)|^2 \\ \vdots & \ddots & \vdots \\ |\mathscr{F}[h_1](|\Omega|)|^2 & \cdots & |\mathscr{F}[h_J](|\Omega|)|^2 \end{bmatrix} \tag{76}$$

Typically, $J < |\Omega|$, making the system (74) under-determined. The unconstrained minimum-norm estimate of the PSD $\lambda^2\check{\Psi}$ can be computed as

$$\widehat{\mathbf{P}} = \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{\dagger}\mathbf{V}, \tag{77}$$

where $\dagger$ denotes the pseudoinverse. A typical constraint that must be enforced onto $\widehat{\mathbf{P}}$ is its nonnegativity, since this is not automatically guaranteed by the pseudoinverse. An iterative system such as the following one can be used to this end:

$$\begin{aligned} \mathbf{R}_k &= \mathbf{V} - \mathbf{F}^T\widehat{\mathbf{P}}_k, \\ \widehat{\mathbf{P}}_{k+1} &= \left|\widehat{\mathbf{P}}_k + \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{\dagger}\mathbf{R}_k\right|. \end{aligned} \tag{78}$$

This system can be modified so to enforce further constraints or priors on $\widehat{\mathbf{P}}_{k+1}$, such as its smoothness or sparsity with respect to some representation [Daubechies et al., 2004], or the symmetries of the Fourier-domain PSDs.

As bank of filters, one can use the Fourier transform of a smaller size, block transforms such as the discrete cosine transform, wavelets, etc. Figure 19 shows two examples of the variances (71) with respect to the Fourier transform of size $35 \times 35$, one estimated for a Samsung S5K2L2 CMOS sensor from the dataset of Sect. 3 and one for a Foveon X3 sensor at ISO 6400 of a Sigma DP1 Merrill camera. While the latter PSD shows a nonuniformity representing colored noise, the former is virtually flat (i.e., white noise), with random fluctuations merely due to finite sample set. Larger size PSDs, estimated via the recursive method (78), are shown in Fig. 20. Both in Fig. 19 and in Fig. 20, we indicate the power level of an ideal flat spectrum of i.i.d. standard white noise (i.e., $35^2 = 1225$ and $255^2 = 65025$ for a $35\times35$ and $255\times255$ Fourier transform, respectively, as can be computed from (7) or (51)). The PSDs estimated from the Samsung sensor accurately match this level, since they are approximately flat and were estimated from standardized errors with $\lambda = 1$ (69); the match is very good also for the Foveon X3 PSDs, since the nonuniformity is anyway quite mild and localized along one line.
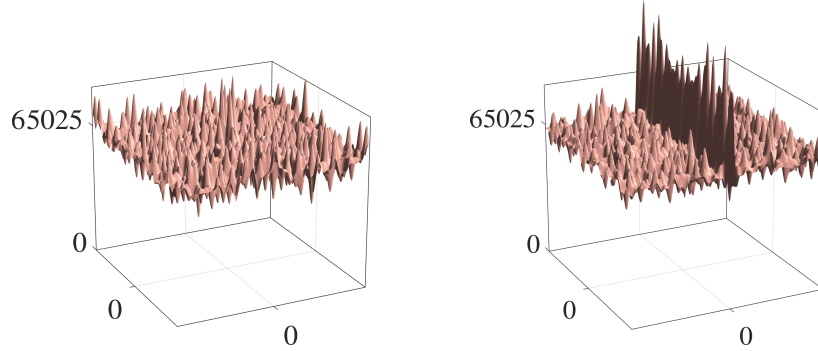
Fig. 20: Estimates of the 255×255 Fourier-domain PSDs computed through the iterative system (78) from the 35×35 measurements shown in Fig. 19.

## 7 Photo-Response Nonuniformity

When a pixel array is exposed to a uniform light source, every pixel element is expected to yield the same underlying signal. However, physical imperfections of the detector elements, such as slight discrepancies in the pixel size and substrate material, cause a deviation from the expected true signal output. This deviation is defined as the photo-response non uniformity (PRNU).

In some cases, the PRNU represents a relevant portion of the image degradation, and thus the image degradation model presented in (26) is no longer adequate. The variable $y$ previously defined as the true signal becomes a random variable dependent on the detector element. Let us define this new variable as $u$. Then,

$$z(x) = u(x) + \sigma(u(x))\xi(x),$$
$$u(x) = \mathcal{N}\left(y(x), cy^2(x)\right). \tag{79}$$

Above we model the physical discrepancies of the detector elements as a Gaussian distribution, with variance depending on the underlying signal $y$ and on a scaling factor $c$. Although (79) describes the nonuniformity as a random process, $u$ is a systematic error which is identically realized at every acquisition (i.e., at each frame), as opposed to $\xi$ which changes randomly at every pixel and at every new capture.

Note that the errors caused by the PRNU are multiplicative, following the same model as described in (30). The univariate standard deviation function described by (27) becomes

$$\mathrm{std}\{z \mid y\} = \sqrt{cy^2 + ay + b}, \tag{80}$$

which is a consequence of treating $z$ as a mixture variate and of $\sigma^2$ being affine. Here $a$, $b$, and $c$ can be estimated using the methods detailed in Sects. 5.1 and 5.2 of this chapter, but considering a second-order polynomial for the noise variance curve estimation.

In general, the PRNU is more evident in images with low gain $a$ or with large exposure time, i.e., large $y(x)$.

# 8 Conclusions

All acquisition systems are affected to some degree by noise. To successfully analyze and process any acquired noisy data, it is necessary to adopt an adequate noise model. This chapter presented several basic noise models that can be used within various imaging applications. First, we introduced white noise models characterized by a flat noise power spectrum. White noise models include the classic signal-independent AWGN noise, as well as various signal-dependent noise models, which are formalized by means of one-parameter families of distributions. Then, we introduced colored noise models, in which the noise power spectral density is not flat, implying that the noise affecting a pixel is correlated with the noise affecting neighboring pixels. An overview of the leading approaches for estimating the noise parameters sufficient to characterize the model was also provided. Overall, these models and methods are suitable for processing imagery from a wide range of acquisition devices, such as digital consumer cameras, X-ray systems, microscopes, telescopes, etc. They especially play a fundamental role in processing pipelines for image restoration (see, e.g., [Foi, 2009; Azzari and Foi, 2016, 2017; Borges et al., 2017, 2018]).

# Acknowledgement

# References

Amer, A. and Dubois, E. (2005). Fast and reliable structure-oriented video noise estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):113–118.

Azzari, L. and Foi, A. (2014a). Gaussian-Cauchy mixture modeling for robust signal-dependent noise estimation. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5357–5361.

Azzari, L. and Foi, A. (2014b). Indirect Estimation of Signal-Dependent Noise With Nonadaptive Heterogeneous Samples. *IEEE Transactions on Image Processing*,

23(8):3459–3467.

Azzari, L. and Foi, A. (2016). Variance Stabilization for Noisy+Estimate Combination in Iterative Poisson Denoising. *IEEE Signal Processing Letters*, 23(8):1086–1090.

Azzari, L. and Foi, A. (2017). Variance stabilization in poisson image deblurring. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 728–731. IEEE.

Borges, L. R., Azzari, L., Bakic, P. R., Maidment, A. D., Vieira, M. A., and Foi, A. (2018). Restoration of low-dose digital breast tomosynthesis. *Measurement Science and Technology*, 29(6):064003.

Borges, L. R., Guerrero, I., Bakic, P. R., Foi, A., Maidment, A. D. A., and Vieira, M. A. C. (2017). Method for simulating dose reduction in digital breast tomosynthesis. *IEEE Transactions on Medical Imaging*, 36(11):2331–2342.

Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J.-B., and Salamero, J. (2010). Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE transactions on medical imaging*, 29(2):442–454.

Cohen, C. A. (1991). *Truncated and Censored Samples*. CRC Press, Boca Raton.

Daubechies, I., Defrise, M., and De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457.

Faà Di Bruno, F. (1857). Note sur une nouvelle formule de calcul différentiel. *Quart. Jour. of Pure and App. Math*, 1:359–360.

Foi, A. (2009). Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12):2609–2629.

Foi, A., Trimeche, M., Katkovnik, V., and Egiazarian, K. (2008). Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754.

Gravel, P., Beaudoin, G., and De Guise, J. A. (2004). A method for modeling noise in medical images. *IEEE Transactions on Medical Imaging*, 23(10):1221–1232.

Greene, W. H. (2000). *Econometric analysis (International edition)*. Pearson US Imports & PHIPEs.

Jähne, B. (2004). *Practical Handbook on Image Processing for Scientific and Technical Applications, Second Edition*. CRC Press, Inc., Boca Raton, FL, USA.

Johnson, J. B. (1928). Thermal agitation of electricity in conductors. *Physical Review*, 32:97–109.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous multivariate distributions, models and applications*, volume 1. New York: John Wiley & Sons.

Lee, J. S. (1981). Refined filtering of image noise using local statistics. *Computer graphics and image processing*, 15(4):380–389.

Lee, J. S. and Hoppel, K. (1989). Noise modeling and estimation of remotely-sensed images. *Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th International Canadian Symposium on Remote Sensing.*, 2:1005–1008.

Liu, C., Freeman, W. T., Szeliski, R., and Kang, S. B. (2006). Noise estimation from a single image. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 901–908. IEEE.

Mäkitalo, M. and Foi, A. (2014). Noise Parameter Mismatch in Variance Stabilization, With an Application to Poisson-Gaussian Noise Estimation. *IEEE Transactions on Image Processing*, 23(12):5348–5359.

Mandel, L. (1959). Fluctuations of Photon Beams: The Distribution of the Photo-Electrons. *Proceedings of the Physical Society*, 74(3):233.

Mastin, G. A. (1985). Adaptive filters for digital image noise smoothing: An evaluation. *Computer Vision, Graphics, and Image Processing*, 31(1):103–121.

Meer, P., Jolion, J. M., and Rosenfeld, A. (1990). A fast parallel algorithm for blind estimation of noise variance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):216–223.

Nyquist, H. (1928). Thermal agitation of electric charge in conductors. *Physical Review*, 32:110–113.

Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes, Third Edition*. Tata McGraw-Hill Education.

Pyatykh, S., Hesser, J., and Zheng, L. (2013). Image Noise Level Estimation by Principal Component Analysis. *IEEE Transactions on Image Processing*, 22(2):687–699.

Starck, J.-L., Murtagh, F. D., and Bijaoui, A. (1998). *Image processing and data analysis: the multiscale approach*. Cambridge University Press.

Sutour, C., Deledalle, C. A., and Aujol, J. F. (2015). Estimation of the Noise Level Function Based on a Nonparametric Detection of Homogeneous Image Regions. *SIAM Journal on Imaging Sciences*, 8(4):2622–2661.

Tijms, H. (2007). *Understanding Probability: Chance Rules in Everyday Life, Second Edition*. Cambridge University Press Cambridge.