

# PoGaIN Supplementary Material

Nicolas Bähler, Majed El Helou, Étienne Objois, Kaan Okumuş, and Sabine Süsstrunk, *Fellow, IEEE*.

## I. MAXIMUM LIKELIHOOD DERIVATION

### A. Poisson-Noise Modeling

Let us denote the observed noisy image as  $y$  and the ground-truth noise-free image as  $x$ . Then, the Poisson-Gaussian model takes the form of the following equation

$$y = \frac{1}{a}\alpha + \beta, \quad \alpha \sim \mathcal{P}(ax), \quad \beta \sim \mathcal{N}(0, b^2). \quad (1)$$

Using the linearity property of expectation, we can compute the expected value

$$\mathbb{E}[y] = \frac{1}{a}\mathbb{E}[\alpha] = \frac{1}{a}ax = x. \quad (2)$$

Further, the variance has the following expression

$$\mathbb{V}[y] = \mathbb{E}\left[\left(\frac{1}{a}\alpha + \beta\right)^2\right] - x^2 = \frac{1}{a^2}\mathbb{E}[\alpha^2] + b^2 - x^2. \quad (3)$$

Given that  $\mathbb{E}[\alpha^2] = ax + a^2x^2$ , we have

$$\mathbb{V}[y] = \frac{x}{a} + x^2 + b^2 - x^2 = \frac{x}{a} + b^2. \quad (4)$$

### B. Likelihood Function of Single-Pixel Image

From the definition of the probability mass function (PMF) of a Poisson random variable  $\alpha$ , we get

$$\mathbb{P}[\alpha = k] = \frac{e^{-ax}(ax)^k}{k!}, \quad k \geq 0. \quad (5)$$

From the relation between the probability density function (PDF) and the PMF of discrete random variable established with the Dirac delta function, i.e.  $f_X(t) = \sum_{k \in \mathbb{Z}} \mathbb{P}[X = k]\delta(t - k)$ , we can derive that

$$f_\alpha(t|a, x) = \sum_{k=0}^{\infty} \frac{e^{-ax}(ax)^k}{k!} \delta(t - k). \quad (6)$$

Let us define  $\alpha' = \frac{1}{a}\alpha$ . Then, the cumulative distribution function (CDF) of this random variable  $\alpha'$  has the following form

$$F_{\alpha'}(t) = \mathbb{P}[\alpha' \leq t] = \mathbb{P}[\alpha \leq at] = F_\alpha(at). \quad (7)$$

By taking the derivative of Equation (7), the PDF of  $\alpha'$  can be found

$$f_{\alpha'}(t) = \frac{dF_{\alpha'}(t)}{dt} = \frac{dF_\alpha(at)}{dt} = af_\alpha(at). \quad (8)$$

Hence, by combining Equations (6) and (8), the likelihood function of  $\alpha'$ , which consists of the first part of the noise model, can be derived

$$\begin{aligned} f_{\alpha'}(t|a, x) &= a \sum_{k=0}^{\infty} \frac{e^{-ax}(ax)^k}{k!} \underbrace{\delta(at - k)}_{=\frac{1}{a}\delta(t - \frac{k}{a})} \\ &= \sum_{k=0}^{\infty} \frac{e^{-ax}(ax)^k}{k!} \delta(t - k/a). \end{aligned} \quad (9)$$

On the other hand, the likelihood function of a Gaussian random variable  $\beta$  with 0 mean is defined as

$$f_\beta(t|b) = \frac{1}{b\sqrt{2\pi}} e^{-t^2/2b^2}. \quad (10)$$

We then combine those equations and find the likelihood function of  $y$ . Since we know that  $\alpha'$  and  $\beta$  are independent of each other, we have that

$$\begin{aligned} \mathcal{L}(y|a, b, x) &= (f_{\alpha'} * f_\beta)(y|a, b, x) \\ &= \sum_{k=0}^{\infty} \frac{(ax)^k}{k!b\sqrt{2\pi}} \exp\left(-ax - \frac{(y - k/a)^2}{2b^2}\right). \end{aligned} \quad (11)$$

### C. Maximum Likelihood Solution for Single-Pixel Image

As derived, the maximum likelihood solution for a single-pixel image is the following

$$\begin{aligned} \hat{a}, \hat{b} &= \arg \max_{a, b} \mathcal{L}(y|a, b, x) \\ &= \arg \max_{a, b} \sum_{k=0}^{\infty} \frac{(ax)^k}{k!b\sqrt{2\pi}} \exp\left(-ax - \frac{(y - k/a)^2}{2b^2}\right). \end{aligned} \quad (12)$$

### D. Likelihood Function of Multi-Pixel Image

We represent images as vectors of pixels, like  $y_n$  and  $x_n$  where  $n \in \mathbb{N}$  is the index of single pixels. Hence, using this notation we obtain

$$\mathcal{L}(y_n|a, b, x_n) = \sum_{k=0}^{\infty} \frac{(ax_n)^k}{k!b\sqrt{2\pi}} \exp\left(-ax_n - \frac{(y_n - k/a)^2}{2b^2}\right). \quad (13)$$

Given  $x$ , i.e., the vector of all  $x_n$ , we can see that  $y_n$  and  $y_{n'}$  are independent  $\forall n \neq n'$ . Therefore, we have

$$\begin{aligned} \mathcal{L}(y|a, b, x) &= \prod_n \sum_{k=0}^{\infty} \frac{(ax_n)^k}{k!b\sqrt{2\pi}} \\ &\quad \exp\left(-ax_n - \frac{(y_n - k/a)^2}{2b^2}\right). \end{aligned} \quad (14)$$

### E. Maximum Likelihood Solution for Multi-Pixel Image

Lastly, we get the following maximization problem

$$\hat{a}, \hat{b} = \arg \max_{a,b} \prod_n \sum_{k=0}^{\infty} \frac{(ax_n)^k}{k!b\sqrt{2\pi}} \exp\left(-ax_n - \frac{(y_n - k/a)^2}{2b^2}\right). \quad (15)$$

Using the strict monotonicity of the logarithm, we can simplify the optimization problem while not altering its results by using the log-likelihood  $\mathcal{LL}$

$$\mathcal{LL}(y|a, b, x) = \sum_n \log \left( \sum_{k=0}^{\infty} \frac{(ax_n)^k}{k!b\sqrt{2\pi}} \exp\left(-ax_n - \frac{(y_n - k/a)^2}{2b^2}\right) \right). \quad (16)$$

Thus, the optimization problem becomes

$$\hat{a}, \hat{b} = \arg \max_{a,b} \mathcal{LL}(y|a, b, x). \quad (17)$$

In order to decrease the high computational complexity, we limit the range of  $k$  to a maximum value  $k_{max}$  which has to be chosen large enough to get a good approximation

$$\hat{a}, \hat{b} \approx \arg \max_{a,b} \sum_n \log \left( \sum_{k=0}^{k_{max}} \frac{(ax_n)^k}{k!b\sqrt{2\pi}} \exp\left(-ax_n - \frac{(y_n - k/a)^2}{2b^2}\right) \right). \quad (18)$$

With bigger values of  $k$  the log-likelihood starts to plateau and does not grow significantly anymore. Hence, by limiting the sum to a large enough  $k_{max}$ , the approximation of the log-likelihood is still good. Typically, we choose  $k_{max} = 100$ . We illustrate this property in the next Figure 1 where we can see how the log-likelihood is indeed reaching a plateau. We average over 25 pixels that we sample randomly, 25 linearly spaced values for  $a \in [1, 100]$  and  $b \in [0.01, 0.15]$ . Additionally, we show the growing computation time needed to obtain those results.

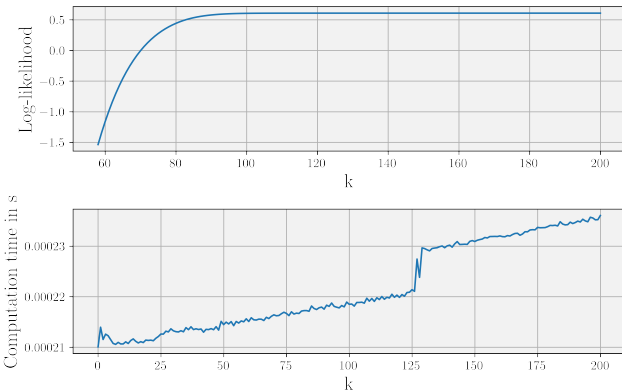


Fig. 1. The evolution of the log-likelihood with bigger  $k$  alongside the computation time.

## II. CUMULANTS

### A. The cumulant of a distribution

For a random variable  $X$  following the distribution  $\mathcal{X}$ , we consider the cumulant-generating function defined as

$$K_{\mathcal{X}}(t) = \log(\mathbb{E}[e^{Xt}]). \quad (19)$$

Then, we define  $\kappa_r[\mathcal{X}]$ , the  $r$ -th cumulant of  $\mathcal{X}$ , as

$$\kappa_r[\mathcal{X}] := K_{\mathcal{X}}^{(r)}(0), \quad (20)$$

with  $K_{\mathcal{X}}^{(r)}(0)$  being the  $r$ -th derivative of  $K_{\mathcal{X}}$  evaluated in 0.

### B. Linearity

The cumulant-generating function of a sum of independent distributions is the sum of their cumulant-generating functions.

*Proof.*

$$\begin{aligned} K_{\mathcal{X}+\mathcal{Y}}(t) &= \log(\mathbb{E}(e^{(X+Y)t})) \\ &= \log(\mathbb{E}[e^{Xt+Yt}]) \\ &= \log(\mathbb{E}[e^{Xt}e^{Yt}]) \\ &= \log(\mathbb{E}[e^{Xt}]\mathbb{E}[e^{Yt}]) \\ &= \log(\mathbb{E}[e^{Xt}]) + \log(\mathbb{E}[e^{Yt}]) \\ &= K_{\mathcal{X}}(t) + K_{\mathcal{Y}}(t). \end{aligned} \quad (21)$$

### C. Homogeneity

The  $r$ -th cumulant is homogeneous of degree  $r$ .

*Proof.*

$$\kappa_r[a\mathcal{X}] = a^r \kappa_r[\mathcal{X}]. \quad (22)$$

### D. Unbiased estimator

For a vector  $x$  obtained by sampling independently  $n$  times from the distribution  $\mathcal{X}$ , the authors of [?] describe an unbiased estimator of  $\kappa_2[\mathcal{X}]$ ,  $\kappa_3[\mathcal{X}]$ ,

$$\kappa_2[\mathcal{X}] = \frac{n}{n-1} m_2(x), \quad \kappa_3[\mathcal{X}] = \frac{n^2}{(n-1)(n-2)} m_3(x), \quad (23)$$

with  $m_2$  being the sample variance (2-rd sample central moment) and  $m_3$  the 3-rd sample central moment, that can be calculated using the formulae taken from [?]

$$\begin{aligned} m_2(x) &= \frac{n-1}{n} \sum_i (x_i - \bar{x})^2 \\ m_3(x) &= \frac{(n-1)(n-2)}{n^2} \sum_i (x_i - \bar{x})^3. \end{aligned} \quad (24)$$

### E. Cumulant of Poisson-Gaussian Noise Model

We have that  $\mathcal{Y} = \frac{\mathcal{P}(a\mathcal{X})}{a} + \mathcal{N}(0, b^2)$  and we want to express  $\kappa_2[\mathcal{Y}]$  and  $\kappa_3[\mathcal{Y}]$  as a function of  $a$  and  $b$ . First, we use Equation (21), and get that,  $\kappa_r[\mathcal{Y}] = \kappa_r\left[\frac{\mathcal{P}(a\mathcal{X})}{a}\right] + \kappa_r[\mathcal{N}(0, b^2)]$ .

1) *Gaussian noise component*: The cumulants of  $\mathcal{N}(0, b^2)$  are known to be

$$\begin{aligned}\kappa_2[\mathcal{N}(0, b^2)] &= b^2 \\ \kappa_3[\mathcal{N}(0, b^2)] &= 0.\end{aligned}\quad (25)$$

2) *Poisson noise component*: Instead of trying to find the cumulant of  $\frac{\mathcal{P}(a\mathcal{X})}{a}$ , we can use Equation (22), and find the cumulant of  $Z \sim \mathcal{Z} = \mathcal{P}(a\mathcal{X})$

$$e^{K_{\mathcal{Z}}(t)} = \sum_k \mathbb{P}[Z = k] e^{tk}. \quad (26)$$

Moreover, we know that

$$\begin{aligned}\mathbb{P}[Z = k] &= \sum_i \mathbb{P}[X = x_i] \mathbb{P}[Z = k | X = i] \\ &= \sum_i n_i \frac{(ax_i)^k e^{-ax_i}}{k!},\end{aligned}\quad (27)$$

where  $n_i = \frac{|\{j: x_j = x_i\}|}{n}$  is the proportion of intensities that are equal to a given one  $x_i$ .

Thus, we have that

$$\begin{aligned}e^{K_{\mathcal{Z}}(t)} &= \sum_k \mathbb{P}[Z = k] e^{tk} \\ &= \sum_k \sum_i n_i \frac{(ax_i)^k e^{-ax_i}}{k!} \exp(t)^k \\ &= \sum_i n_i \frac{e^{-ax_i}}{\exp(-ax_i e^t)} \sum_k \frac{(ax_i e^t)^k \exp(-ax_i e^t)}{k!} \\ &= \sum_i n_i \exp(ax_i(e^t - 1)).\end{aligned}\quad (28)$$

If we further note that,  $f : t \mapsto \sum_i n_i \exp(ax_i(e^t - 1))$ , then, we get that  $K_{\mathcal{Z}}(t) = \log(f(t))$ . Hence, we can now compute the different derivatives of  $K_{\mathcal{Z}}(t)$

$$\begin{aligned}K_{\mathcal{Z}}(t) &= \log(f(t)) \\ K_{\mathcal{Z}}^1(t) &= \frac{f^{(1)}(t)}{f(t)} \\ K_{\mathcal{Z}}^2(t) &= \frac{f^{(2)}(t)f(t) - f^{(1)}(t)^2}{f(t)^2} \\ K_{\mathcal{Z}}^3(t) &= \frac{f(t)[f(t)f^{(3)}(t) - 3f^{(2)}(t)f^{(1)}(t)] + 2f^{(1)}(t)^3}{f(t)^3}.\end{aligned}\quad (29)$$

Further, by evaluating those at 0, we get

$$\begin{aligned}\kappa_0[\mathcal{Z}] &= 0 \\ \kappa_1[\mathcal{Z}] &= a\bar{x} \\ \kappa_2[\mathcal{Z}] &= a\bar{x} + a^2\bar{x}^2 - a^2\bar{x}^2 \\ \kappa_3[\mathcal{Z}] &= a^3[\bar{x}^3 - 3\bar{x}^2\bar{x} + 2\bar{x}^3] + a^2[3\bar{x}^2 - 3\bar{x}^2] + a\bar{x},\end{aligned}\quad (30)$$

using the properties that

$$\begin{aligned}f(0) &= 1 \\ f^{(1)}(0) &= a\bar{x} \\ f^{(2)}(0) &= a\bar{x} + a^2\bar{x}^2 \\ f^{(3)}(0) &= a\bar{x} + 3a^2\bar{x}^2 + 2a^3\bar{x}^3.\end{aligned}\quad (31)$$

Then, using Equation (22), we obtain

$$\begin{aligned}\kappa_2\left[\frac{\mathcal{P}(a\mathcal{X})}{a}\right] &= \frac{\bar{x}}{a} + \bar{x}^2 - \bar{x}^2 \\ \kappa_3\left[\frac{\mathcal{P}(a\mathcal{X})}{a}\right] &= \bar{x}^3 - 3\bar{x}^2\bar{x} + 2\bar{x}^3 + 3\frac{\bar{x}^2}{a} - 3\frac{\bar{x}^2}{a} + \frac{\bar{x}}{a^2}.\end{aligned}\quad (32)$$

3) *Poisson-Gaussian Noise Model*: By putting Equations (25) and (32) together, we obtain the complete expression of the cumulants

$$\begin{aligned}\kappa_2[\mathcal{Y}] &= \frac{\bar{x}}{a} + \bar{x}^2 - \bar{x}^2 + b^2 \\ \kappa_3[\mathcal{Y}] &= \bar{x}^3 - 3\bar{x}^2\bar{x} + 2\bar{x}^3 + 3\frac{\bar{x}^2}{a} - 3\frac{\bar{x}^2}{a} + \frac{\bar{x}}{a^2}.\end{aligned}\quad (33)$$

### III. CNN ARCHITECTURE

The detailed architecture of the CNN can be found in table I.

TABLE I  
ARCHITECTURE OF THE CNN

Layer	Out channels	Parameters
Input	1	-
Conv2D	16	kernel_size = (3, 3), padding = same
ReLU	16	-
BatchNorm	16	over the channels
MaxPool2D	16	pool_size = (2, 2)
Conv2D	32	kernel_size = (3, 3), padding = same
ReLU	32	-
BatchNorm	32	over the channels
MaxPool2D	32	pool_size = (2, 2)
Conv2D	64	kernel_size = (3, 3), padding = same
ReLU	64	-
BatchNorm	64	over the channels
MaxPool2D	64	pool_size = (2, 2)
Dense	16	-
ReLU	16	-
BatchNorm	16	over the channels
Dropout	16	rate = 0.5
Dense	4	-
ReLU	4	-
Dense	2	-
Linear	2	-