Unbiased Injection of Signal-Dependent Noise in Variance-Stabilized Range

Lucas Rodrigues Borges, Marcelo Andrade da Costa Vieira, and Alessandro Foi

***Abstract*—The design, optimization, and validation of many image processing or image-based analysis systems often requires testing of the system performance over a dataset of images corrupted by noise at different signal-to-noise ratio regimes. A noise-free ground-truth image may not be available, and different SNRs are simulated by injecting extra noise into an already noisy image. However, noise in real-world systems is typically signal-dependent, with variance determined by the noise-free image. Thus, also the noise to be injected shall depend on the unknown ground-truth image. To circumvent this issue, we consider the additive injection of noise in variance-stabilized range, where no previous knowledge of the ground-truth signal is necessary. Specifically, we design a special noise-injection operator that prevents the errors on expectation and variance that would otherwise arise when standard variance-stabilizing transformations are used for this task. Thus, the proposed operator is suitable for accurately injecting signal-dependent noise even to images acquired at very low counts.**

***Index Terms*—Noise injection, variance stabilization, optimization, Anscombe transformation, Poisson noise.**

I. Introduction

Many image processing and image-based analysis systems are sensitive to variations on the image quality. It is imperative to understand the effect that image SNR may exert on the performance of these methods. Therefore it is a common approach to evaluate them over sets of images corrupted by different noise levels.

For instance, the optimization of the radiation dose in X-ray systems is of great importance in medical applications. Acquiring several images of the same patient at different SNR regimes is prohibitive due to radiation-related risks. In this case, it is common to perform pre-clinical trials by simulating different radiation doses through injection of noise into a noisy standard-dose image [1]–[8].

In many applications, the observations are corrupted by signal-dependent non-additive errors related to the inherent uncertainties such as photon accumulation, which is often modeled through the Poisson distribution. The data distribution hence depends on its expectation, and simulations thus require an underlying noise-free signal, or ground-truth image. However, this poses a challenge, as such noise-free image may not be available. One approach consists in estimating the true signal, either using local statistics from a single observation [2] or through previous modeling of the acquisition system [5]. However, estimating the underlying signal is not always feasible, since it may introduce errors due to smearing and blurring through smoothing, as pointed by the authors themselves [2], [5]. Another approach consists in creating a signal-independent noise mask with the desired noise properties and installing dependency by assuming that the observed Poisson variable is a good approximation of the true signal [1], [3], [4], [6], which provides good performance only at high-count regimes.

In previous works we proposed the injection of additive Gaussian noise in a variance-stabilized range, resulting in an approximately Poissonian distribution in the signal range [7], [8]. In this way, no previous knowledge of the true signal is needed, and the noise injection is performed in a pixel-by-pixel manner, avoiding potential problems due to smoothing or due to inaccurate estimation of the signal. However, due to the use of generic variance-stabilizing transformations (ideally meant for the denoising task), the method [7], [8] is biased for small counting and performs poorly in applications where the counting rate is reduced.

Thus, here we introduce an operator built using a pair of forward and inverse transformations jointly optimized specifically for the noise-injection task. Such operator is capable of performing unbiased noise injection even at very low counting rates (≈ 1 count per pixel), allowing, *e.g.*, the accurate simulation of reduced-dose X-ray images from inputs that are already in photon-limited imaging modalities. To the best of our knowledge, this is the first work that addresses this problem specifically taking into account the low-count regime, which is the most challenging case for two fundamental reasons. Firstly, at low counts the SNR of the input image is already very low, making it impossible to obtain a trustworthy estimate of the underlying noise-free image through denoising. Even state-of-the-art filters [9] invariably introduce smearing, blurring, or other artifacts [10], [11]. Secondly, variance-stabilization techniques are typically designed based on large-count asymptotics, which is well known to lead to inaccurate stabilization and bias at low counts [12], [13]. In particular, we demonstrate our methodology for the Poisson observation case, which is arguably the most challenging one since exact variance stabilization of Poisson data is known to be fundamentally impossible to achieve [14].

II. Preliminaries

A. Poisson observation model and problem formulation

Consider $\mathring{z}_i, i = 1, 2, \ldots, N$ as the acquired pixels forming an image $\mathring{z}$. We model each $\mathring{z}_i$ as an independent scaled Poisson

This work was supported by the Academy of Finland (project no. 252547), by the São Paulo Research Foundation (FAPESP grant no. 2013/18915-5), and by the Brazilian Foundation for the Coordination of Improvement of Higher Education Personnel (CAPES grant no. 88881.030443/2013-01).

L.R. Borges and M.A.C. Vieira are with the University of São Paulo, Brazil (e-mail: lucas.rodrigues.borges@usp.br, mvieira@sc.usp.br); A. Foi is with Tampere University of Technology, Finland (e-mail: alessandro.foi@tut.fi).

random variable with underlying signal $y_i \geq 0$,

$$\mathring{z}_i = \mathring{\pi}_i / \mathring{\lambda} \qquad \mathring{\pi}_i \sim \mathcal{P}(\mathring{\lambda} y_i) \tag{1}$$

where $\mathrm{E}\{\mathring{z}_i|y_i\} = y_i$, $\mathrm{var}\{\mathring{z}_i|y_i\} = y_i/\mathring{\lambda}$, and $\mathring{\lambda} > 0$ is a scaling factor that controls the relative noise strength.

Our goal is to obtain from $\mathring{z}_i$ a new set of *noisier* scaled Poisson observations z_i with a smaller scaling factor $\lambda < \mathring{\lambda}$:

$$z_i = \pi_i / \lambda \qquad \pi_i \sim \mathcal{P}(\lambda y_i). \tag{2}$$

Thus z_i has the same mean as $\mathring{z}_i$ but larger standard deviation:

$$\mathrm{E}\{z_i|y_i\} = y_i = \mathrm{E}\{\mathring{z}_i|y_i\}, \quad \mathrm{std}\{z_i|y_i\} = \sqrt{\frac{y_i}{\lambda}} = \frac{\mathrm{std}\{\mathring{z}_i|y_i\}}{\sqrt{\lambda/\mathring{\lambda}}}. \tag{3}$$

A definition such as (2) assumes knowledge of the noise-free y_i, which means that obtaining z_i from $\mathring{z}_i$ may be possible only subject to certain approximations. Pragmatically, we formulate an optimization problem where the sought solution is a non-negative $z_i \geq 0$ that solves (3) in a least-squares sense.

B. Variance-stabilizing transformation

Variance-stabilizing transformations (VSTs) are commonly applied to cease the dependency between the noise-free signal and the noise variance, allowing, *e.g.*, suppression of signal-dependent noise through many off-the-shelf filters for additive models [13]. The typical VST for Poisson data is the Anscombe transformation, which for observations (1) takes the form

$$a_{\mathring{\lambda}}(\mathring{z}_i) = 2\sqrt{\mathring{\lambda}\mathring{z}_i + \frac{3}{8}}, \tag{4}$$

yielding stabilized variables that can be treated as normal with unit variance: $a_{\mathring{\lambda}}(\mathring{z}_i) \sim \mathcal{N}\left(a_{\mathring{\lambda}}(y_i) + 1/8, 1\right)$ [15].

III. Noise injection

A. Definition

Let f be the VST for $\mathring{z}$. We consider addition of further noise to the stabilized $f(\mathring{z})$ and hence we aim at constructing an operator $\Phi_{\lambda,\mathring{\lambda}}$ of the form

$$z_i = \Phi_{\lambda,\mathring{\lambda}}(\mathring{z}_i) = g\left(f(\mathring{z}_i) + \sigma n_i\right) \tag{5}$$

where $n_i \sim \mathcal{N}(0,1)$ and $\sigma \geq 0$ so that $\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}_i)$ satisfies (3). For instance, a tedious but otherwise simple analysis based on mixtures of noncentral χ^2 distributions [16] shows that setting

$$f = a_{\mathring{\lambda}}, \quad \sigma^2 = \frac{\mathring{\lambda}}{\lambda} - 1, \quad g(x) = \frac{1}{4\mathring{\lambda}}\left(x^2 - \frac{3}{2} - \sigma^2\right) \tag{6}$$

yields

$$E\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}_i)\middle|y_i\right\} = y_i, \tag{7}$$

$$\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}_i)\middle|y_i\right\} = \frac{y_i}{\lambda} + \frac{\mathring{\lambda}^2 + \mathring{\lambda}\lambda - 2\lambda^2}{8\mathring{\lambda}^2\lambda^2}. \tag{8}$$

We wish to emphasize that the equalities (7) and (8) are not based on asymptotics or other approximations and are thus precise for any combination of $y_i \geq 0$ and $\mathring{\lambda} \geq \lambda \geq 0$. In particular, (7) shows that (5) and (6) solve the left-hand side of (3) exactly for any $y \geq 0$. When this happens we say that $\Phi_{\lambda,\mathring{\lambda}}$ performs an (exact) *unbiased noise injection in variance-stabilized range*. Equation (8) shows that the right-hand side of (3) is solved only approximately, with good relative accuracy for large y. However, this approximation of the desired variance may be rather coarse for small λ or small y. Moreover, (6) does not account for the non-negativity requirement, yielding variables that can be as low as $g(0) < 0$. The clipped $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z}) = \max\left\{0, \Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\right\}$ ensures non-negativity, but impacts both (7) and (8), particularly at low count rates [17]. Figure 1 (left) illustrates how the expectation and standard deviation of $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ differ from the goal (3) at low counts. Let us note that (6) is *per se* already an improvement over [7], [8], where the inverse transformation was an inverse of the Anscombe transformation designed to be unbiased *with* filtering and *without* noise injection, which led to a systematic extra positive bias ($\approx \lambda^{-1}/4$) at both large and small counts.

B. Optimization

The operator $\Phi^+_{\lambda,\mathring{\lambda}}$ as defined by (6) fulfills (3) asymptotically. To improve the operator at low counts, we define a practical optimization task. Following [18], we model f and g through rational-polynomial functions,

$$f(x) = 2\sqrt{\frac{\sum_{k=0}^{J} p_k x^k}{\sum_{k=0}^{K} q_k x^k}} = 2\sqrt{\frac{P(x)}{Q(x)}}, \tag{9}$$

$$g(x) = \frac{\sum_{k=0}^{N} r_k x^k}{\sum_{k=0}^{M} s_k x^k} = \frac{R(x)}{S(x)}, \tag{10}$$

and optimize the coefficients of the polynomials P, Q, R, and S under the constraint that f and g approach the definitions (6) asymptotically, which ensures [19] that also (3) are attained asymptotically through (7) and (8). Thus we require

$$\frac{P(x)}{Q(x)} - \mathring{\lambda}x - \frac{3}{8} \underset{x\to+\infty}{\longrightarrow} 0, \tag{11}$$

$$\frac{R(x)}{S(x)} - \frac{1}{4\mathring{\lambda}}\left(x^2 - \frac{3}{2} - \sigma^2\right) \underset{x\to+\infty}{\longrightarrow} 0. \tag{12}$$

These force $K = J-1$ and $M = N-2$. For simplicity, we here set $J = 3$ and $N = 4$, for which (11) and (12) hold provided

$$p_3 = \mathring{\lambda} q_2, \qquad q_1 = p_2 - \frac{3p_3}{8}, \qquad r_4 = \frac{3 + 2\sigma^2}{\mathring{\lambda} s_2},$$

$$r_3 = \frac{\mathring{\lambda} s_1}{3 + 2\sigma^2}, \qquad r_2 = \frac{\mathring{\lambda} s_0}{3 + 2\sigma^2} - \frac{3 + 2\sigma^2}{2\mathring{\lambda}}.$$

We can always fix $q_0 = s_0 = 1$, hence (9) and (10) depend only on few free parameters $\Pi = [\, p_3\ p_2\ p_1\ p_0\ r_4\ r_3\ r_1\ r_0\,]$, which we can vary as to minimize the discrepancy errors

$$\mathcal{E}_{\mathrm{E}}(y) = \mathrm{E}\left\{\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} - y$$

$$\mathcal{E}_{\mathrm{std}}(y) = \mathrm{std}\left\{\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} - \sqrt{\frac{y}{\lambda}}$$

between the actual expectation and standard deviation of $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ and their desired values (3). In particular, we define the optimal parameters Π^* as the solution of

$$\Pi^* = \underset{\Pi}{\mathrm{argmin}} \int_0^{+\infty} \frac{\mathcal{E}^2_{\mathrm{E}}(y)}{y_\epsilon^2} + \frac{\mathcal{E}^2_{\mathrm{std}}(y)}{y_\epsilon}\, \mathrm{d}y, \tag{13}$$

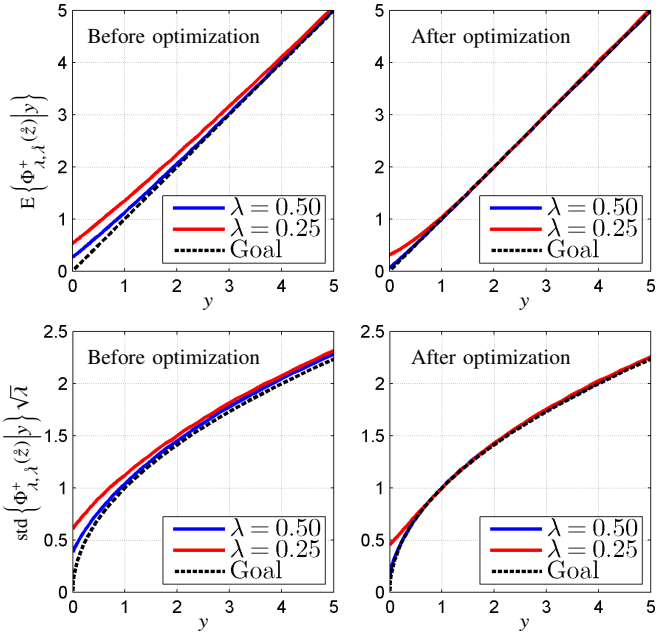

Figure 1. Expectation and standard deviation of $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ before (left) and after (right) optimization ($\lambda = 0.25, 0.50$, $\mathring{\lambda} = 1$).

where $y_\epsilon = \max\{\epsilon, y\}$, $\epsilon > 0$. The integral cost (13) controls the relative quadratic errors; the integral is always finite, at $y \to 0^+$ by using a fixed positive ϵ, and at $y \to +\infty$ due to the constraints on the polynomial coefficients.

Note that $\Pi_{\text{init}} = \left[\begin{array}{cccccccc} 0 & 0 & 1 & 3/8 & 0 & 0 & -1/2 & 0 \end{array}\right]$ gives exactly (6).

IV. Experiments

To solve (13) we employ a multi-start Nelder-Mead direct-search method [20], [21], with initial condition given by Π_{init} and set $\epsilon = 0.01$. Because cases with $\mathring{\lambda} \neq 1$ can be treated using $\frac{1}{\mathring{\lambda}}\Phi^+_{\lambda/\mathring{\lambda},1}\left(\mathring{\lambda}\mathring{z}\right)$, without loss of generality we always assume $\mathring{\lambda} = 1$. We consider two illustrative cases $\lambda = 0.50$ and $\lambda = 0.25$, which respectively correspond to a half and to a quarter exposure (or dose in case of X-ray radiation). Figure 2 shows the optimized functions f and g, where the latter is plotted as $g^+ = \max\{0, g\}$ to incorporate the non-negativity of $\Phi^+_{\lambda,\mathring{\lambda}}$. Note the irregular profile, particularly for g with $\lambda = 0.50$, typical of optimized VSTs (see, *e.g.*, [18], [22]). Figure 1 (right) presents the expectation and standard deviation of the optimized $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$, which approximate the goal (3) in practice perfectly for y as low as 1. A comparison with Figure 1 (left) demonstrates the significant improvement over the unoptimized operator (6).

In Figure 3, we compare the cumulative distribution function (CDF) and mean of z according to the ideal target Poisson distribution (2) with those obtained from $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ before and after optimization, for a few combinations of λ, and y. The first two sub-plots from the left ($\lambda = 0.50$ and $\lambda = 0.25$) show that a decrease on λ increases the error on the mean, which is corrected by the optimization. The rightmost sub-plot ($y = 20$) shows that the method converges for high values of the signal even before the optimization is performed.

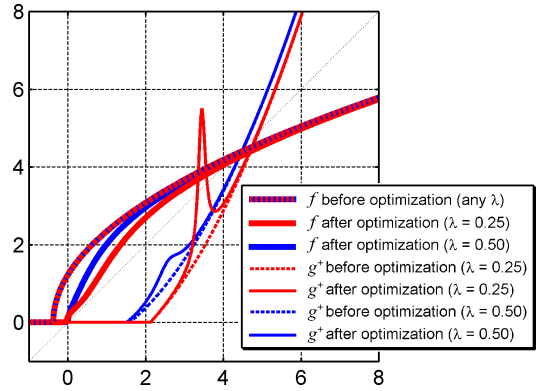

Figure 2. Functions f and $g^+ = \max\{0, g\}$ before and after optimization of $\Phi^+_{\lambda,\mathring{\lambda}}$ ($\lambda = 0.25, 0.50$, $\mathring{\lambda} = 1$).

Further, Figure 4 demonstrates the application of the proposed approach to a low-count image $\mathring{z}$ of fluorescent cells. For this experiment, we have access to a virtually noise-free ground-truth signal y, from which we can obtain a reference ideal noisy z for $\lambda = 0.5$ according to (2), against which we can compare the other noisy images. Firstly, we can notice the bias in the unoptimized $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ solution (6), particularly in the background, which is visibly brighter. It is instead difficult to detect qualitative differences between the optimized $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ (13) obtained from $\mathring{z}$ and the ideal z obtained from y: despite the already low SNR of $\mathring{z}$, many thin filament structures and localized features can be still recognized. We also wish to compare the result of our method to an image ζ, $\lambda\zeta_i \sim \mathcal{P}(\lambda\hat{y}_i)$, generated from an estimate $\hat{y}$ of y by a state-of-the-art denoising method [9], [13] applied to $\mathring{z}$. We can observe a loss of detail and contrast when $\hat{y}$ is used as an approximation of y. For instance, the filament at bottom-left of the image is lost through filtering, and contours are overall less sharp in ζ than in any of the $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ images. We wish to emphasize that differences between ζ ans z are caused *only* by the differences between $\hat{y}$ and y, since $\mathcal{P}$ used the same pseudo-random pattern for all images.

As a final experiment, we compare the accuracy of the proposed method to that of the direct injection of signal-dependent noise with $\pi_i \approx \mathcal{N}\left(\mathring{z}_i\lambda, \mathring{z}_i\lambda\left(1 - \lambda/\mathring{\lambda}\right)\right)$, as used by, *e.g.*, [1], [3], [4], [6]. This normal distribution is actually an approximation of the binomial distribution $\mathcal{B}\left(\mathring{z}_i\mathring{\lambda}, \lambda/\mathring{\lambda}\right) = \mathcal{P}(\lambda y_i)$ which, while formally ideal, is never used directly because it does not generalize well to the non-pure Poisson scenario typical of practical applications. Similar to $\Phi^+_{\lambda,\mathring{\lambda}}$, also in this case we need to enforce non-negativity which similarly impacts the expectation and variance leading to unwanted bias, as illustrated in Figure 5 for $\lambda = 0.25$. The case $\lambda = 0.50$, omitted due to space limitation, demonstrates comparably negligible expectation bias for the two methods, and a minor deficiency in the directly injected variance.

V. Discussion and Conclusions

We have proposed an operator capable of changing the relative strength of signal-dependent noise corrupting an image. The injection is pixelwise and is performed in a variance-

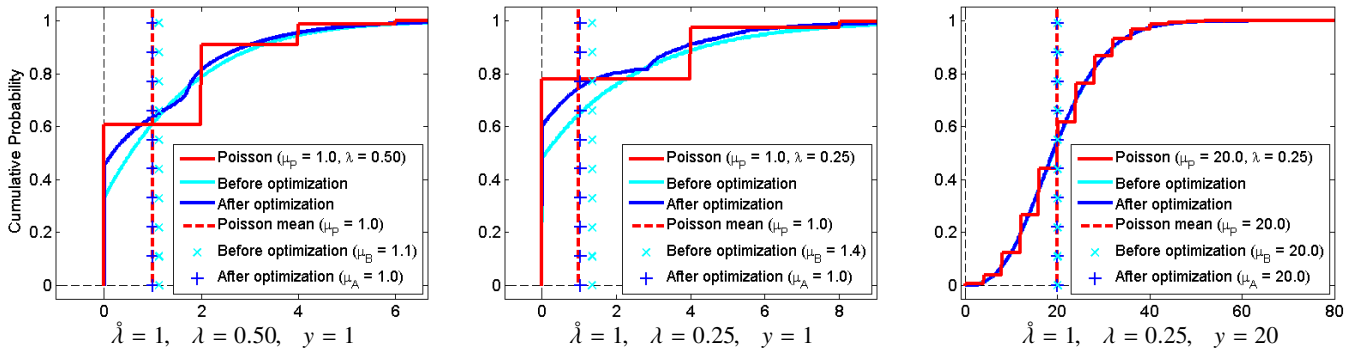

Figure 3. Cumulative distribution function (CDF) and mean μ_P of z according to the ideal target Poisson distribution (2) vs the CDFs and means μ_B, μ_A obtained from $\Phi^+_{\lambda,\mathring{\lambda}}(\mathring{z})$ before optimization (6) and after optimization (13). Each subplot corresponds to different combinations of y and λ, while $\mathring{\lambda}=1$.

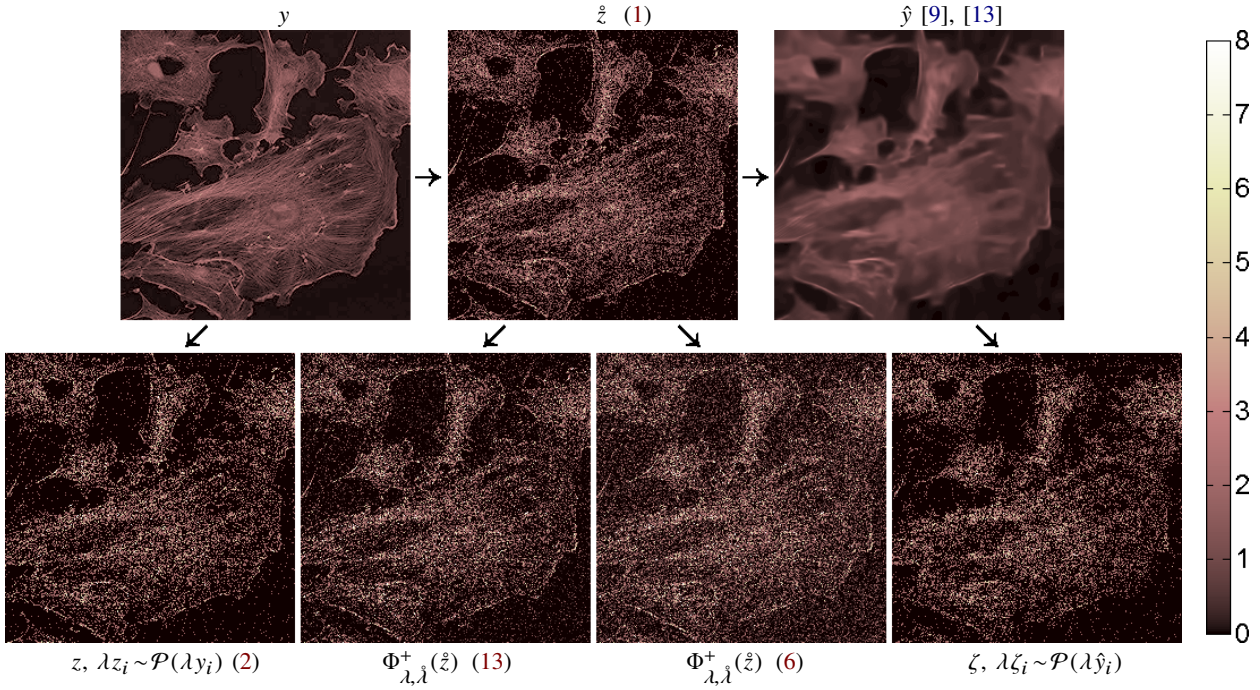

Figure 4. Comparison of various approaches for obtaining a lower-SNR image (3) from a Poisson observation $\mathring{z}$ (1), for $\lambda=0.50$, $\mathring{\lambda}=1$: results by proposed noise-injection before (6) and after optimization (13), Poisson ζ on a denoised estimate $\hat{y}$, ideal reference Poisson z (2) on noise-free ground-truth y.

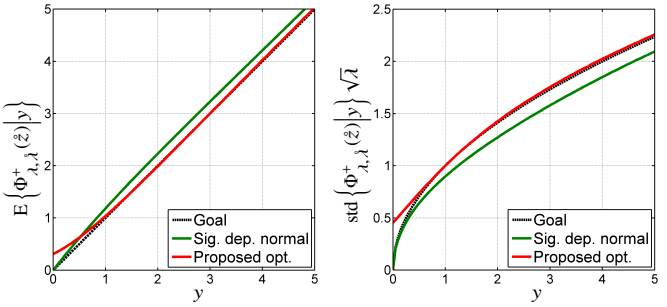

Figure 5. Comparison of expectation and standard deviation of the random variables produced by direct injection of signal-dependent noise to the image intensities vs the proposed method of injection in variance-stabilized range with optimized transformations (λ = 0.25, $\mathring{\lambda}$ = 1).

stabilized range, where knowledge of the ground-truth image is unnecessary. We showed that the initial form of the operator, based on asymptotics, performs poorly at low counts, and have thus developed an optimized operator that significantly improves the results across the dynamic range. Since this approach is based on univariate mappings and on injection of spatially uncorrelated noise, the total computational cost is proportional to the number of pixels in the image. The present goal (3) and the corresponding cost (13) involve only the first two moments of z, but it is of course possible to include extra higher-order moments, and employ higher-order rational polynomials with $J>3$ and $N>4$ or other parametric families of transformations. One may naturally expect better fit from increasing the number of parameters.

In this letter we have considered the Poisson distribution as the most challenging case. The proposed approach can be easily generalized to other distributions, and we shall consider noise injection also for Poisson-Gaussian combinations, as well as the interaction with quantization. Also, we wish to further investigate the case of spatially correlated noise. Finally, we aim to obtain a closed form of the optimized $\Phi^+_{\lambda,\mathring{\lambda}}$, readily applicable to any $0<\lambda<\mathring{\lambda}$.

REFERENCES

[1] M. Båth, M. Håkansson, A. Tingberg, and L. G. Månsson, "Method of simulating dose reduction for digital radiographic systems," *Radiation protection dosimetry*, vol. 114, no. 1-3, pp. 253–259, 2005. [Online]. Available: http://dx.doi.org/10.1093/rpd/nch540

[2] W. J. Veldkamp, L. J. Kroft, J. P. A. van Delft, and J. Geleijns, "A technique for simulating the effect of dose reduction on image quality in digital chest radiography," *J. Digital Imaging*, vol. 22, no. 2, pp. 114–125, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10278-008-9104-5

[3] A. Svalkvist and M. Båth, "Simulation of dose reduction in tomosynthesis," *Medical Physics*, vol. 37, no. 1, pp. 258–269, 2010. [Online]. Available: http://dx.doi.org/10.1118/1.3273064

[4] A. Mackenzie, D. R. Dance, A. Workman, M. Yip, K. Wells, and K. C. Young, "Conversion of mammographic images to appear with the noise and sharpness characteristics of a different detector and x-ray system," *Medical Physics*, vol. 39, no. 5, pp. 2721–2734, 2012. [Online]. Available: http://dx.doi.org/10.1118/1.4704525

[5] S. Zabic, Q. Wang, T. Morton, and K. M. Brown, "A low dose simulation tool for CT systems with energy integrating detectors," *Medical Physics*, vol. 40, no. 3, 2013. [Online]. Available: http://dx.doi.org/10.1118/1.4789628

[6] A. Mackenzie, D. R. Dance, O. Diaz, and K. C. Young, "Image simulation and a model of noise power spectra across a range of mammographic beam qualities," *Medical Physics*, vol. 41, no. 12, pp. 121 901–121 901, 2014. [Online]. Available: http://dx.doi.org/10.1118/1.4900819

[7] L. R. Borges, H. C. R. de Oliveira, P. F. Nunes, and M. A. C. Vieira, "Method for inserting noise in digital mammography to simulate reduction in radiation dose," in *Proc. SPIE Medical Imaging 2015*, vol. 9412, no. 5J, 2015. [Online]. Available: http://dx.doi.org/10.1117/12.2082257

[8] L. R. Borges, H. C. R. de Oliveira, P. F. Nunes, P. R. Bakic, A. D. A. Maidment, and M. A. C. Vieira, "Method for simulating dose reduction in digital mammography using the Anscombe transformation," *Medical Physics*, vol. 43, no. 6, pp. 2704–2714, 2016. [Online]. Available: http://dx.doi.org/10.1118/1.4948502

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007. [Online]. Available: http://dx.doi.org/10.1109/TIP.2007.901238

[10] E. Vansteenkiste, D. Van der Weken, W. Philips, and E. Kerre, "Perceived image quality measurement of state-of-the-art noise reduction schemes," in *Proc. Advanced Concepts for Intelligent Vision Systems (ACIVS2006)*. Springer, 2006, pp. 114–126. [Online]. Available: http://dx.doi.org/10.1007/11864349_11

[11] P. Chatterjee and P. Milanfar, "Bias modeling for image denoising," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 856–859. [Online]. Available: http://dx.doi.org/10.1109/ACSSC.2009.5469988

[12] J.-L. Starck, F. D. Murtagh, and A. Bijaoui, *Image processing and data analysis: the multiscale approach*. Cambridge University Press, 1998. [Online]. Available: http://www.multiresolution.com/cupbook.pdf

[13] M. Mäkitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, Jan 2011. [Online]. Available: http://dx.doi.org/TIP.2010.2056693

[14] J. H. Curtiss, "On transformations used in the analysis of variance," *Ann. Math. Statist.*, vol. 14, no. 2, pp. 107–122, 06 1943. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177731452

[15] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948. [Online]. Available: http://www.jstor.org/stable/2332343

[16] L. R. Borges, M. A. C. Vieira, and A. Foi, Supplementary to the article "Unbiased Injection of Signal-Dependent Noise in Variance-Stabilized Range", 2016.

[17] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.sigpro.2009.04.035

[18] M. Mäkitalo and A. Foi, "Noise parameter mismatch in variance stabilization, with an application to Poisson-Gaussian noise estimation," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5348–5359, Dec 2014. [Online]. Available: http://dx.doi.org/10.1109/TIP.2014.2363735

[19] S. K. Bar-Lev and P. Enis, "On the construction of classes of variance stabilizing transformations," *Statistics & Probability Letters*, vol. 10, no. 2, pp. 95–100, 1990. [Online]. Available: http://dx.doi.org/10.1016/0167-7152(90)90002-O

[20] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965. [Online]. Available: http://dx.doi.org/10.1093/comjnl/7.4.308

[21] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*. SIAM, 2009, vol. 8. [Online]. Available: http://dx.doi.org/10.1137/1.9780898718768

[22] A. Foi, "Optimization of variance-stabilizing transformations," 2009, preprint. [Online]. Available: http://www.cs.tut.fi/~foi/papers/Foi-OptimizationVST-preprint.pdf

Supplementary to the article "Unbiased Injection of Signal-Dependent Noise in Variance-Stabilized Range"

Lucas Rodrigues Borges, Marcelo Andrade da Costa Vieira, and Alessandro Foi

In this appendix we detail the mathematical steps that lead from (5) and (6) to (7) and (8) on the second page of the article.

We begin from expanding the output of the noise-injection operator as

$$\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}) = \frac{x^2}{4\mathring{\lambda}} - \frac{3}{8\mathring{\lambda}} - \frac{\sigma^2}{4\mathring{\lambda}},$$

where x is obtained by applying f to $\mathring{z}$ followed by the addition of noise,

$$x = 2\sqrt{\mathring{\lambda}\mathring{z} + \frac{3}{8}} + \sigma n, \qquad n(\cdot) \sim \mathcal{N}(0,1).$$

Throughout our analysis, we treat $\left\{x^2 \middle| y\right\}$ as a mixture distribution with mixture components $C_j = \left\{x^2 \middle| \mathring{z} = \mathring{\lambda}^{-1} j\right\}$ and mixture weights w_j equal to the scaled Poisson probability $\mathrm{P}\left(\mathring{z} = \mathring{\lambda}^{-1} j \middle| y\right)$, $j = 0, 1, 2, \ldots$. According to this mixture model, we have

$$\mathrm{E}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}) \middle| y\right\} = \frac{\mathrm{E}\left\{x^2 \middle| y\right\}}{4\mathring{\lambda}} - \frac{3}{8\mathring{\lambda}} - \frac{\sigma^2}{4\mathring{\lambda}} = \frac{1}{4\mathring{\lambda}} \sum_{j=0}^{n} w_j m_j - \frac{3}{8\mathring{\lambda}} - \frac{\sigma^2}{4\mathring{\lambda}} \tag{14}$$

$$\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z}) \middle| y\right\} = \frac{\mathrm{var}\left\{x^2 \middle| y\right\}}{16\mathring{\lambda}^2} = \frac{1}{16\mathring{\lambda}^2}\left(\sum_{j=0}^{n} w_j \left(m_j^2 + s_j^2\right) - m^2\right) \tag{15}$$

where m_j and s_j^2 are respectively the mean and variance of C_j, and $m = \sum_{j=0}^{n} w_j m_j = \mathrm{E}\left\{x^2 \middle| y\right\}$.

For any given value of $\mathring{z}$, the conditional distribution of x is a normal centered at $f(\mathring{z})$:

$$\{x | \mathring{z}\} \sim \mathcal{N}\left(2\sqrt{\mathring{\lambda}\mathring{z} + \frac{3}{8}}, \sigma^2\right).$$

Hence,

$$\left\{\frac{x}{\sigma} \middle| \mathring{z}\right\} \sim \mathcal{N}\left(\frac{2}{\sigma}\sqrt{\mathring{\lambda}\mathring{z} + \frac{3}{8}}, 1\right)$$

Therefore, for any given value of $\mathring{z}$, x^2/σ^2 follows a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter $\mu^2 = \mathrm{E}^2\left\{\frac{x}{\sigma} \middle| \mathring{z}\right\}$. The conditional expectation and variance are thus

$$\mathrm{E}\left\{\frac{x^2}{\sigma^2} \middle| \mathring{z}\right\} = 1 + \mu^2 = 1 + \frac{4}{\sigma^2}\left(\mathring{\lambda}\mathring{z} + \frac{3}{8}\right),$$
$$\mathrm{var}\left\{\frac{x^2}{\sigma^2} \middle| \mathring{z}\right\} = 2 + 4\mu^2 = 2 + \frac{16}{\sigma^2}\left(\mathring{\lambda}\mathring{z} + \frac{3}{8}\right).$$

Hence,

$$\mathrm{E}\left\{x^2 \middle| \mathring{z} = \mathring{\lambda}^{-1} j\right\} = m_j = \sigma^2 + 4\left(j + \frac{3}{8}\right), \tag{16}$$

$$\mathrm{var}\left\{x^2 \middle| \mathring{z} = \mathring{\lambda}^{-1} j\right\} = s_j^2 = 2\sigma^4 + 16\sigma^2\left(j + \frac{3}{8}\right). \tag{17}$$

Substituting (16) into (14) and recalling that $\mathring{\lambda}\mathring{z} \sim \mathcal{P}\left(\mathring{\lambda}y\right)$ yields (7):

$$\begin{aligned}\mathrm{E}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} &= \frac{1}{4\mathring{\lambda}}\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)\left(\sigma^2+4\left(j+\frac{3}{8}\right)\right)-\frac{3}{8\mathring{\lambda}}-\frac{\sigma^2}{4\mathring{\lambda}} = \\ &= \frac{1}{\mathring{\lambda}}\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)j = \frac{1}{\mathring{\lambda}}\sum_{j=0}^{n}\mathrm{P}\left(\mathring{\lambda}\mathring{z}=j\middle|y\right)j = \frac{1}{\mathring{\lambda}}\mathrm{E}\left\{\mathring{\lambda}\mathring{z}\middle|y\right\} = y. \end{aligned} \tag{18}$$

Combining (18) with (14) also gives $m = 4\mathring{\lambda}y + \frac{3}{2} + \sigma^2$. We can now substitute (17) into (15):

$$\begin{aligned}\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} &= \frac{1}{16\mathring{\lambda}^2}\left(\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)\left(\left(\sigma^2+4\left(j+\frac{3}{8}\right)\right)^2+2\sigma^4+16\sigma^2\left(j+\frac{3}{8}\right)\right)-\left(4\mathring{\lambda}y+\frac{3}{2}+\sigma^2\right)^2\right) = \\ &= \frac{1}{16\mathring{\lambda}^2}\left(\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)\left(16j^2+24j\sigma^2+12j-16y^2\mathring{\lambda}^2-8y\sigma^2\mathring{\lambda}-12y\mathring{\lambda}+2\sigma^4+6\sigma^2\right)\right).\end{aligned}$$

Noting, as in (18), that $\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)j = \mathrm{E}\left\{\mathring{\lambda}\mathring{z}\middle|y\right\} = \mathring{\lambda}y$, we obtain

$$\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} = \frac{1}{\mathring{\lambda}^2}\left(\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)\left(j^2+\mathring{\lambda}y\sigma^2-y^2\mathring{\lambda}^2+\frac{\sigma^4}{8}+\frac{3\sigma^2}{8}\right)\right).$$

Next, since $y^2\mathring{\lambda}^2 = \left(\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)j\right)^2$, we have $\sum_{j=0}^{n}\mathrm{P}\left(\mathring{\lambda}\mathring{z}=j\middle|y\right)j^2 - y^2\mathring{\lambda}^2 = \mathrm{var}\left\{\mathring{\lambda}\mathring{z}\middle|y\right\} = \mathring{\lambda}y$ and hence

$$\begin{aligned}\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} &= \frac{1}{\mathring{\lambda}^2}\left(\sum_{j=0}^{n}\mathrm{P}\left(\mathring{z}=\mathring{\lambda}^{-1}j\middle|y\right)\left(\mathring{\lambda}y\left(1+\sigma^2\right)+\frac{\sigma^4}{8}+\frac{3\sigma^2}{8}\right)\right) = \\ &= \frac{y\left(1+\sigma^2\right)}{\mathring{\lambda}}+\frac{\sigma^4}{\mathring{\lambda}^2 8}+\frac{3\sigma^2}{8\mathring{\lambda}^2}.\end{aligned}$$

Since $\sigma^2 = \frac{\mathring{\lambda}}{\lambda} - 1$, we finally obtain (8):

$$\mathrm{var}\left\{\Phi_{\lambda,\mathring{\lambda}}(\mathring{z})\middle|y\right\} = \frac{y}{\lambda}+\frac{\mathring{\lambda}^2+\mathring{\lambda}\lambda-2\lambda^2}{8\mathring{\lambda}^2\lambda^2}.$$