Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise

Markku Mäkitalo and Alessandro Foi

***Abstract*—Many digital imaging devices operate by successive photon-to-electron, electron-to-voltage, and voltage-to-digit conversion. This process is subject to various signal-dependent errors, which are typically modelled as Poisson-Gaussian noise. The removal of such noise can be approached indirectly by applying a variance-stabilizing transformation (VST) to the noisy data, denoising the stabilized data with a Gaussian denoising algorithm, and finally applying an inverse VST to the denoised data. The generalized Anscombe transformation (GAT) is often used for variance stabilization, but its unbiased inverse transformation has not been rigorously studied in the past. We introduce the exact unbiased inverse of the GAT and show that it plays an integral part in ensuring accurate denoising results. We demonstrate that this exact inverse leads to state-of-the-art results without any notable increase in the computational complexity compared to the other inverses. We also show that this inverse is optimal in the sense that it can be interpreted as a maximum likelihood inverse. Moreover, we thoroughly analyze the behaviour of the proposed inverse, which also enables us to derive a closed-form approximation for it. This paper generalizes our work on the exact unbiased inverse of the Anscombe transformation, which we have presented earlier for the removal of pure Poisson noise.**

***Index Terms*—denoising, photon-limited imaging, Poisson-Gaussian noise, variance stabilization.**

I. Introduction

Many imaging devices, such as digital cameras and any device equipped with a CCD or a CMOS sensor, capture images by successive photon-to-electron, electron-to-voltage, and voltage-to-digit conversion. This capturing process is subject to various signal-dependent errors, and a standard way to model these errors is to consider them as Poisson-Gaussian noise. Specifically, photon emission and sensing are inherently random physical processes, which in turn substantially contribute to the randomness in the sensor output. Thus, the noise model employs a Poisson component in order to account for this signal-dependent uncertainty. Complementarily, the additive Gaussian component accounts for the other signal-independent noise sources involved in the capturing chain, such as thermal noise. This modelling has been successfully used in several practical applications, e.g., in noise fitting and denoising of clipped and non-clipped raw CCD data [1], [2], in the denoising of fluorescence microscopy images [3], [4], in fluorescent-spot detection [3], and in astronomy [6].

There are two main options for approaching the problem of denoising images corrupted by signal-dependent noise. One option is to directly consider the statistics of the particular noise model, and take advantage of these properties and observations in designing an effective denoising algorithm. For the case of Poisson-Gaussian noise, this avenue has been explored in, e.g., [4]. Alternatively, the problem can be tackled in a modular fashion through variance stabilization. This general denoising process involves three steps. First, the noisy data is modified by applying a nonlinear variance-stabilizing transformation (VST) specifically designed for the chosen noise model. Note that in an image corrupted by signal-dependent noise, the noise variance is typically not constant and varies with the expectation of the pixel value. For instance, the variance of a Poisson variable equals its mean. Thus, the rationale of applying a VST is to remove this signal-dependency by rendering the noise variance constant throughout the image. In particular, the transformed data can be assumed to have an approximately Gaussian noise distribution with a known constant (e.g., unitary) variance. Hence, the second step is to treat the noisy data with any algorithm designed for the removal of Gaussian noise. Finally, the desired estimate of the unknown noise-free image is obtained by applying an inverse VST to the denoised data (i.e., by returning the denoised data to the original range). This modular approach has several practical advantages: Not only is the problem of Gaussian noise removal a well studied and widely covered topic, with a plethora of denoising algorithms to choose from, but it also allows the practical implementation of the denoising framework to be divided into self-contained modules, which may be designed and optimized independent of each other.

In the case of Poisson-Gaussian noise, the generalized Anscombe transformation (GAT) [6] is commonly used for stabilizing the noise variance. This transformation generalizes the classical Anscombe transformation [9], which was designed for the pure Poisson case. Even though the GAT is a well-known transformation, its corresponding exact unbiased inverse transformation has been neglected in the past. We introduce the exact unbiased inverse of the GAT and show that it plays an integral part in ensuring accurate denoising results. We also conclude that the poor denoising performance shown in earlier works (especially in the low-intensity cases, where an asymptotically unbiased inverse is particularly inaccurate by its nature) is mostly due to applying an unsuitable inverse transformation, and not simply due to the inability of the GAT to stabilize the noise variance adequately.

This work was supported by the Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006–2011, project no. 252547, Academy Research Fellow 2011–2016, and project no. 129118, Postdoctoral Researcher's Project 2009–2011) and by Tampere Doctoral Programme in Information Science and Engineering (TISE).

This paper extends the authors' preliminary conference publication [5], by both elaborating on the theoretical aspects and widening the scope of the experimental consideration.

The authors are with the Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland (e-mail: firstname.lastname@tut.fi).

To the best of our knowledge, this paper is the first rigorous study of the exact unbiased inverse of the GAT for Poisson-Gaussian noise. Importantly, our results are exact and successfully applicable also for finite parameter values, while previous works [6] have considered only the large-parameter asymptotic case. Moreover, our experimental results demonstrate that our approach leads to state-of-the-art denoising results, when the exact unbiased inverse is combined with a state-of-the-art Gaussian denoising algorithm. We also show that this inverse is optimal in the sense that it can be interpreted as a maximum likelihood inverse under certain reasonable assumptions. Further, we provide a thorough analysis of the behaviour of the proposed inverse. In particular, we show that the exact unbiased inverse of the GAT can be approximated with great accuracy by adding a simple correction term to the exact unbiased inverse of the Anscombe transformation, which we introduced in [7]. This also enables us to derive a closed-form approximation for it, by adding the same correction term to the closed-form approximation [8] of the exact unbiased inverse of the Anscombe transformation. Finally, we provide asymptotical, as well as global integral and global supremum error bounds for both of these approximations.

Overall, this paper generalizes our earlier work [7], [8], in which we presented an exact unbiased inverse of the Anscombe transformation [9] and similarly showed its importance in guaranteeing accurate denoising results for the case of pure Poisson noise. It is also worth noting that as a consequence of this conceptual similarity, employing the exact unbiased inverse remains computationally very inexpensive also in this generalization. In other words, the proposed approach does not introduce any notable computational overhead; the choice of the Gaussian denoising algorithm completely dominates the execution time. These observations are consistent with the discussion on computational complexity in [7].

The rest of the paper is organized as follows: Section II presents some preliminaries about Poisson-Gaussian noise and about stabilizing its variance with the GAT. In Section III, which is the core of our contribution, we discuss how to construct an exact unbiased inverse of the GAT, the optimality and asymptotic behaviour of this inverse, and how to approximate it with a closed-form expression. Section IV consists of various experiments. In particular, first we examine the case of denoising with known noise parameters, after which we consider the case with unknown parameter values. Then we inspect how the denoising performance changes, when the ratio between the Poisson and Gaussian noise component changes. We conclude the section with a brief commentary on the computational complexity of the proposed method. Finally, in Section V we discuss the obtained results. For the sake of readability, most of the mathematical derivations and proofs are presented in the Appendix.

II. Preliminaries

A. Poisson-Gaussian noise

Let $\overset{*}{z}_i, i = 1, \ldots, N$, be the observed pixel values obtained through an image acquisition device. We model each $\overset{*}{z}_i$ as an independent random Poisson variable p_i with an underlying mean value y_i, scaled by $\alpha > 0$ and corrupted by additive Gaussian noise $\overset{*}{n}_i$ of mean μ and standard deviation $\overset{*}{\sigma}$. In other words,

$$\overset{*}{z}_i = \alpha p_i + \overset{*}{n}_i, \tag{1}$$

where $p_i \sim \mathcal{P}(y_i)$ and $\overset{*}{n}_i \sim \mathcal{N}(\mu, \overset{*}{\sigma}^2)$. Thus, we can define Poisson-Gaussian noise as

$$\eta_i = \overset{*}{z}_i - \alpha y_i. \tag{2}$$

The problem of denoising an image corrupted by Poisson-Gaussian noise is then equivalent to estimating the underlying noise-free image y given the noisy observations $\overset{*}{z}$. For clarity, we note that the overhead asterisk in $\overset{*}{z}$, $\overset{*}{\sigma}$ and $\overset{*}{n}_i$ is used to distinguish these variables before and after the affine transformations (4), which we will employ in order to reduce the number of parameters to be considered; the rest of the paper will only deal with the corresponding transformed variables z, σ and n_i, and with y and p_i, unless noted otherwise.

B. Variance stabilization with the generalized Anscombe transformation

Assuming $\overset{*}{z}$ is distributed according to (1), we can apply the generalized Anscombe transformation [6]

$$f(\overset{*}{z}) = \begin{cases} \frac{2}{\alpha}\sqrt{\alpha\overset{*}{z} + \frac{3}{8}\alpha^2 + \overset{*}{\sigma}^2 - \alpha\mu}, & \overset{*}{z} > -\frac{3}{8}\alpha - \frac{\overset{*}{\sigma}^2}{\alpha} + \mu \\ 0, & \overset{*}{z} \leq -\frac{3}{8}\alpha - \frac{\overset{*}{\sigma}^2}{\alpha} + \mu \end{cases} \tag{3}$$

to $\overset{*}{z}$ in order to (approximately) stabilize its variance to unity, i.e., $\operatorname{var}\{f(\overset{*}{z})|y, \overset{*}{\sigma}\} \approx 1$. Note that for the pure Poisson case (i.e., $\alpha = 1$, $\sigma = 0$, and $\mu = 0$), this coincides with the traditional Anscombe transformation [9] used for stabilizing data corrupted by Poisson noise.

The number of parameters which define the transformation (3) can be reduced significantly by simple variable substitutions

$$z = \frac{\overset{*}{z} - \mu}{\alpha}, \quad \sigma = \frac{\overset{*}{\sigma}}{\alpha}, \tag{4}$$

which affinely map each pixel $\overset{*}{z}_i$ to z_i, a random (non-scaled) Poisson variable p_i corrupted by additive Gaussian noise n_i of mean 0 and standard deviation σ:

$$z_i = p_i + n_i, \tag{5}$$

where $p_i \sim \mathcal{P}(y_i)$ and $n_i \sim \mathcal{N}(0, \sigma^2)$. In particular, the probability distribution of z is

$$p(z \mid y, \sigma) = \sum_{k=0}^{+\infty} \left(\frac{y^k e^{-y}}{k!} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-k)^2}{2\sigma^2}} \right). \tag{6}$$

Thus, according to (3), z can be stabilized with the transformation

$$f_\sigma(z) = \begin{cases} 2\sqrt{z + \frac{3}{8} + \sigma^2}, & z > -\frac{3}{8} - \sigma^2 \\ 0, & z \leq -\frac{3}{8} - \sigma^2 \end{cases}. \tag{7}$$

In other words, for any of the parameters α and μ, we can stabilize the variance of $\overset{*}{z}$ by means of variable substitutions (4), followed by the transformation (7). Then, after applying an inverse transformation $\mathcal{I}$ of (7) to the denoised data D, we simply return to the original range by inverting (4), i.e. setting the final estimate of the expected value of $\overset{*}{z}$ to be $\alpha\mathcal{I}(D) + \mu$. Note that since (4) and its inverse are affine, they

do not introduce any bias in the estimation. Thus, in the rest of the paper, we consider only the stabilization of z, which is the observed data after the variable substitution (4), and by GAT we refer to the corresponding transformation (7).

Figure 1(a) shows the forward transformation (7) for the parameter values $\sigma = 0.01, 1, 2, 3$, and the corresponding standard deviations of the stabilized variables $f_\sigma(z)$ are shown in Figure 1(b). Note that there is a particular overshoot in the standard deviation at around $\sigma = 2$ for low values of y, but it begins to settle down towards the desired value 1 as σ increases. On the other hand, when σ is very small (i.e., when the noise is practically pure Poisson), there is a significant undershoot for low values of y, which is an inherent limitation of the generalized Anscombe transformation. Regardless of which transformation is used, it is not possible to stabilize the variance to unity for $y = 0$, $\sigma = 0$, since in this case z has zero variance.

III. Exact unbiased inverse transformation

A. Definition

Assuming the denoising of $f_\sigma(z)$ was successful, we may treat the denoised data D as the expected value $E\{f_\sigma(z) \mid y, \sigma\}$. However, due to the nonlinearity of f_σ, we generally have

$$f_\sigma^{-1}\left(E\{f_\sigma(z) \mid y\}\right) \neq E\{z \mid y\}. \tag{8}$$

In practice, this means that applying the algebraic inverse f_σ^{-1} to the denoised data will, in general, produce a biased estimate of y. This problem is traditionally mitigated by using an asymptotically unbiased inverse instead of the algebraic inverse. However, as we also show in this work, it is not sufficient for guaranteeing accurate denoising results, in particular for low-intensity images. Thus, there is a need for an exact unbiased inverse that can be used in all situations.

The exact unbiased inverse of the generalized Anscombe transformation (7) is in fact a family of inverse transformations $\mathcal{I}_\sigma$, parametrized by σ, that maps the values $E\{f_\sigma(z) \mid y, \sigma\}$ to the desired values $E\{z \mid y, \sigma\}$:

$$\mathcal{I}_\sigma : E\{f_\sigma(z) \mid y, \sigma\} \longmapsto E\{z \mid y, \sigma\}. \tag{9}$$

Since we trivially know $E\{z \mid y, \sigma\} = y$ for any given y, constructing the inverse requires us to compute the values $E\{f_\sigma(z) \mid y, \sigma\}$, analogously to how the exact unbiased inverse of the Anscombe transformation was computed in [7]. In this more general case, it is computed as

$$E\{f_\sigma(z) \mid y, \sigma\} = \int_{-\infty}^{+\infty} f_\sigma(z)\, p(z \mid y, \sigma)\, dz \tag{10}$$

$$= \int_{-\infty}^{+\infty} 2\sqrt{z + \frac{3}{8} + \sigma^2} \sum_{k=0}^{+\infty} \left(\frac{y^k e^{-y}}{k!\sqrt{2\pi\sigma^2}} e^{-\frac{(z-k)^2}{2\sigma^2}} \right) dz.$$

The exact unbiased inverse transformations $\mathcal{I}_\sigma$ for the parameter values $\sigma = 0.01, 1, 2, 3$ are shown in Figure 1(c), along with the corresponding algebraic inverses of (7) for the comparison. Note that even though visually $\mathcal{I}_\sigma$ resembles a clipped algebraic inverse to some extent, replacing the standard algebraic inverse by its clipped counterpart does not provide any practical improvement.

B. Optimality

Under certain reasonable assumptions, we can show that the exact unbiased inverse $\mathcal{I}_\sigma$ is optimal in the sense that it coincides with a maximum likelihood (ML) inverse. In [7], we made a similar conclusion about the exact unbiased Anscombe inverse for the pure Poisson case, and thus, the discussion below follows naturally the same lines.

Let us assume that instead of obtaining a perfectly accurate denoising result $D = E\{f_\sigma(z) \mid y, \sigma\}$ (which we assumed in the definition of $\mathcal{I}_\sigma$ in Section III-A), the pointwise mean squared error of our estimate D is

$$\varepsilon^2 = E\left\{ \left(D - E\{f_\sigma(z) \mid y, \sigma\} \right)^2 \right\} > 0. \tag{11}$$

In practice the probability density function $p(D \mid y)$ of D is unknown, but for simplicity we assume that it is symmetric and unimodal, with mode at $E\{f_\sigma(z) \mid y, \sigma\}$ and variance ε^2. Note that the actual value of ε^2 has very little to do with the variance of z or $f_\sigma(z)$, since the estimation accuracy of the denoising algorithm depends mainly on the features of the signal (such as the complexity of details and patterns); thus, it is sensible to treat ε^2 as independent of both y and σ.

In other words, we assume

$$\xi = D - E\{f_\sigma(z) | y, \sigma\} \sim U_0, \tag{12}$$

where U_0 is a unimodal distribution with mode at 0 with probability density u_0 (e.g., U_0 could be the normal $\mathcal{N}(0, \varepsilon^2)$). This implies that $u_0\left(\bar{\xi}\right) \leq u_0\left(\breve{\xi}\right)$, if either $\bar{\xi} \leq \breve{\xi} \leq 0$ or $\bar{\xi} \geq \breve{\xi} \geq 0$; in particular $u_0(0) = \max_\xi(u_0)$, and $u_0\left(D - E\{f_\sigma(z)|y, \sigma\}\right) = p(D|y)$. Let us also remark that, for any $\sigma \geq 0$, $E\{f_\sigma(z)|y, \sigma\}$ is a continuous and monotonically increasing function of y.

While (12) also formally implies that D is an unbiased estimate of $E\{f_\sigma(z)|y, \sigma\}$, in fact also unknown estimation-bias errors can be considered as contributors of ε^2, with the symmetry of the distribution about $E\{f_\sigma(z)|y, \sigma\}$ reflecting our uncertainty about the sign of the bias.

By treating D as the data, the ML estimate of y is defined as

$$\mathcal{I}_{\text{ML}}(D) = \arg\max_y p(D \mid y). \tag{13}$$

Under the above assumptions, this equals to (see Appendix B for details)

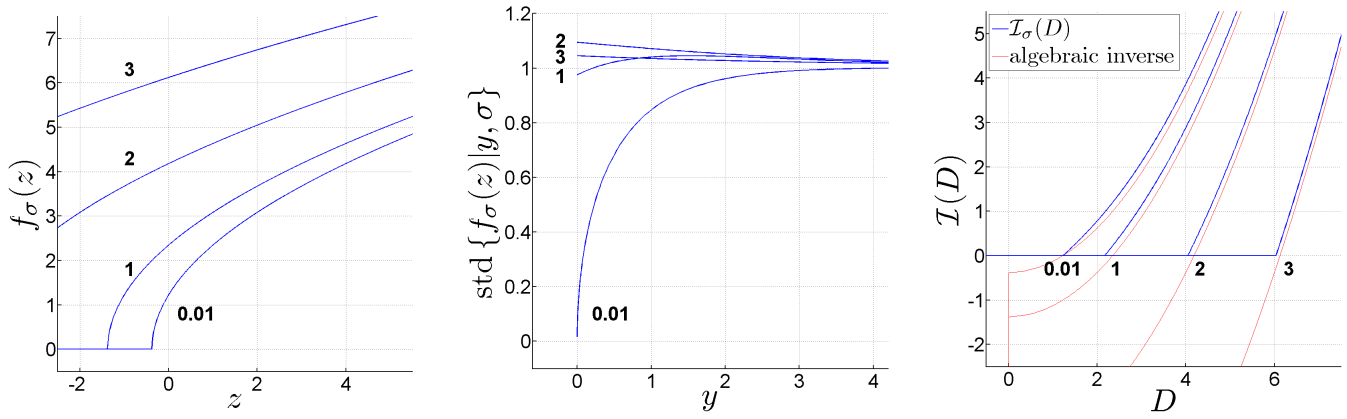
$$\mathcal{I}_{\text{ML}}(D) = \begin{cases} \mathcal{I}_\sigma(D), \text{ if } D \geq E\{f_\sigma(z) \mid 0, \sigma\} \\ 0, \text{ if } D < E\{f_\sigma(z) \mid 0, \sigma\}. \end{cases} \tag{14}$$

Thus, the exact unbiased inverse $\mathcal{I}_\sigma$ coincides with this form of ML inverse.

C. Asymptotic behaviour

As we explicitly construct the inverse mapping (9) only for a finite grid of values, it is also of interest to examine the asymptotic behaviour of $\mathcal{I}_\sigma$. When the standard deviation σ of the Gaussian noise component is large, we may formulate the exact unbiased inverse $\mathcal{I}_\sigma$ in terms of the exact unbiased inverse Anscombe transformation $\mathcal{I}_0$ [7] as

$$\mathcal{I}_\sigma \approx \mathcal{I}_0 - \sigma^2. \tag{15}$$



(a) (b) (c)

Fig. 1. The generalized Anscombe transformation (7) for the parameter values $\sigma = 0.01, 1, 2, 3$. (a) The forward transformations $f_\sigma(z)$, (b) The standard deviations of the stabilized variables $f_\sigma(z)$, (c) The exact unbiased inverse transformations $\mathcal{I}_\sigma$, compared with the corresponding algebraic inverses of (7).

Likewise, when σ is very small, we may do the same approximation. A rigorous derivation of this approximation is provided in Appendix A.

Considering the direct asymptotics with respect to $D = E\{f_\sigma(z)\,|y,\sigma\}$ and σ, we show that for a given fixed σ (and thus considering large y),

$$y = \mathcal{I}_\sigma(D) = \mathcal{I}_0(D) - \sigma^2 + \mathcal{O}\left(D^{-4}\right). \tag{16}$$

For a fixed y (and thus considering large σ), we have $\sigma^2 \sim D^2$ and, consequently,

$$\begin{aligned} y = \mathcal{I}_\sigma(D) &= \mathcal{I}_0(D) - \sigma^2 + \mathcal{O}\left(D^{-2}\right) \\ &= \mathcal{I}_0(D) - \sigma^2 + \mathcal{O}\left(\sigma^{-2}\right). \end{aligned} \tag{17}$$

Finally, for fixed y and small σ, we have

$$y = \mathcal{I}_\sigma(D) = \mathcal{I}_0(D) - \sigma^2 + \mathcal{O}\left(\sigma^2\right). \tag{18}$$

The orders $\mathcal{O}\left(\sigma^{-2}\right)$ and $\mathcal{O}\left(\sigma^2\right)$ of the error terms in (17) and (18), for various fixed values of y, are illustrated in Figure 2. Similarly, Figure 3 visualizes the order $\mathcal{O}\left(D^{-4}\right)$ of the error term in (16). For the detailed derivation of these results, we again refer the reader to Appendix A.

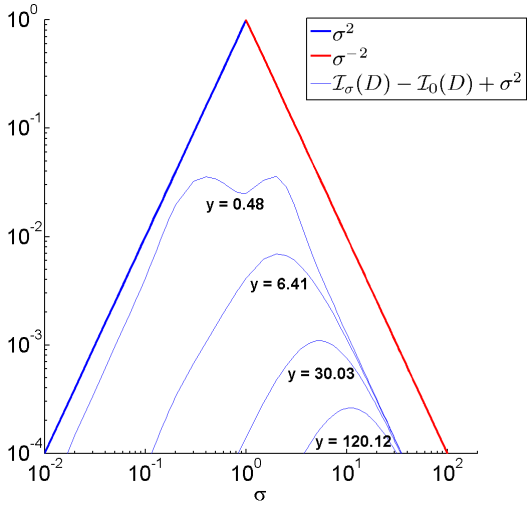


Fig. 2. The error $\mathcal{I}_\sigma(D) - \mathcal{I}_0(D) + \sigma^2 = \mathcal{O}\left(\sigma^2\right)$ as a function of σ for various fixed values of y, compared with σ^{-2} and σ^2, confirming the asymptotics in (17) and (18).

D. Global accuracy

Apart from the above asymptotic results, we studied the global accuracy of the approximation (15) in terms of the variance-normalized integral criterion (weighted L^2 squared), yielding

$$\begin{aligned} &\left\| \frac{\mathcal{I}_0(E\{f_\sigma(z)\,|y,\sigma\}) - \sigma^2 - y}{\mathrm{std}\,\{z|y,\sigma\}} \right\|_2^2 = \\ &= \int_0^{+\infty}\int_0^{+\infty} \frac{\left(\mathcal{I}_0(E\{f_\sigma(z)\,|y,\sigma\}) - \sigma^2 - y\right)^2}{y+\sigma^2}\,dy\,d\sigma = 0.0028 \end{aligned} \tag{19}$$

and in terms of the maximum absolute difference (L^∞), yielding

$$\begin{aligned} &\left\|\mathcal{I}_0(E\{f_\sigma(z)\,|y,\sigma\}) - \sigma^2 - y\right\|_\infty = \\ &\quad = \max_{\sigma\geq 0, y\geq 0} |\mathcal{I}_\sigma(E\{f_\sigma(z)\,|y,\sigma\}) - y| = 0.0470 \end{aligned} \tag{20}$$

with the maximum in (20) attained at $\sigma = 0.4$ and $y = 0$. In Figure 4, the blue surface (below) shows the error $|\mathcal{I}_0(E\{f_\sigma(z)\,|y,\sigma\}) - \sigma^2 - y|$ associated with (15), thus visualizing the overall accuracy and the maximum error of the approximation.

E. Practical implementation

Due to the discussed asymptotic behaviour, and given that $\mathcal{I}_0$ is already available (either in accurate numerical form or as closed-form analytical approximation [7], [8]), to compute $\mathcal{I}_\sigma$ it is sufficient to tabulate $E\{f_\sigma(z)\,|y,\sigma\}$ (10) only for a finite grid of values, and resort to interpolation (between the grid values) and to the asymptotic form $\mathcal{I}_0(D) - \sigma^2$ (outside of the grid). In particular, for our experiments, we considered 96 non-equispaced values $\sigma \in \{0.01, \ldots, 50\}$ and 1199 non-equispaced values of $y \in \{0, \ldots, 200\}$ and calculated $E\{f_\sigma(z)\,|y,\sigma\}$ on such 96×1199 grid. Our Matlab software implementing this inverse transformation is available online at http://www.cs.tut.fi/~foi/invansc.

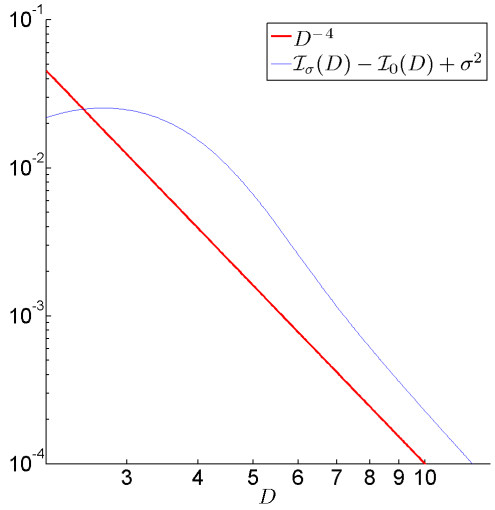


Fig. 3. The error $\mathcal{I}_\sigma(D) - \mathcal{I}_0(D) + \sigma^2 = \mathcal{O}\left(D^{-4}\right)$ as a function of D (for fixed $\sigma = 1$), compared with D^{-4}, confirming the asymptotics in (16).

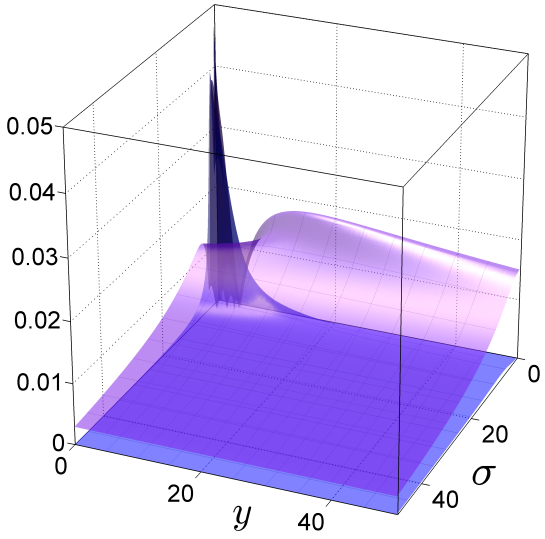


Fig. 4. The error surface $|\mathcal{I}_0(E\{f_\sigma(z)\,|y,\sigma\}) - \sigma^2 - y|$ associated with (15) (blue surface below), and the corresponding surface $|\widetilde{\mathcal{I}}_\sigma(E\{f_\sigma(z)\,|y,\sigma\}) - y|$ associated with the closed-form approximation (21) (purple surface on top).

F. Closed-form approximation

A closed-form approximation of $\mathcal{I}_\sigma$ can be obtained from the closed-form approximation $\widetilde{\mathcal{I}}_0$ [8] of $\mathcal{I}_0$ as

$$\widetilde{\mathcal{I}}_\sigma(D) = \widetilde{\mathcal{I}}_0(D) - \sigma^2 = \tag{21}$$
$$= \frac{1}{4}D^2 + \frac{1}{4}\sqrt{\frac{3}{2}}D^{-1} - \frac{11}{8}D^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}D^{-3} - \frac{1}{8} - \sigma^2.$$

Of course, the negative powers of D become irrelevant when D is large, i.e., when σ or y are large, yielding the asymptotic inverse

$$\mathcal{I}_{\text{asy}}(D) = \frac{1}{4}D^2 - \frac{1}{8} - \sigma^2. \tag{22}$$

As in Section III-D, we studied the accuracy of the approximation (21) in terms of the variance-normalized integral criterion (weighted L^2 squared) and the maximum absolute difference L^∞, with the corresponding results being

$$\left\| \frac{\widetilde{\mathcal{I}}_\sigma(E\{f_\sigma(z)\,|y,\sigma\}) - y}{\text{std}\,\{z|y,\sigma\}} \right\|_2^2 = 0.0069, \tag{23}$$

and

$$\left\| \widetilde{\mathcal{I}}_\sigma(E\{f_\sigma(z)\,|y,\sigma\}) - y \right\|_\infty = 0.0468, \tag{24}$$

where the maximum of (24) is again attained at $\sigma = 0.4$ and $y = 0$. The error surface associated with this criterion is also illustrated in Figure 4 (purple surface on top).

In terms of the impact on denoising quality, the discrepancies between $\widetilde{\mathcal{I}}_\sigma$ and $\mathcal{I}_\sigma$ are very small, as affirmed by (23) and (24), and by the experimental results and observations in the next section. These results and considerations are consistent with those presented in [8].

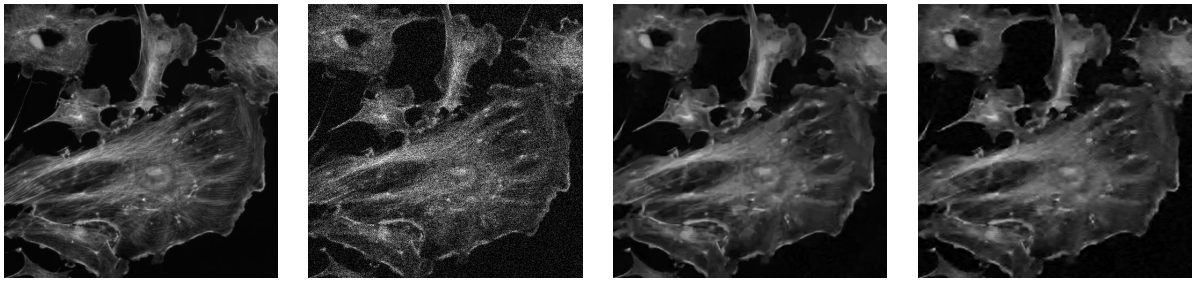
IV. Experiments

A. Denoising with known parameter values α and σ

We evaluate the denoising performance associated with the proposed exact unbiased inverse $\mathcal{I}_\sigma$ by considering the Cameraman (256×256), Fluorescent Cells (512×512), and Lena (512×512) test images. For each image, we scale the original image to eight different peak values (1, 2, 5, 10, 20, 30, 60, 120), and corrupt them with Poisson-Gaussian noise ($\alpha = 1$, $\sigma =$ peak/10) according to (5), as was done in [4]. Let us remark that since $\alpha = 1$, we have also $\overset{*}{\sigma} = \sigma$, as can be seen from (4).

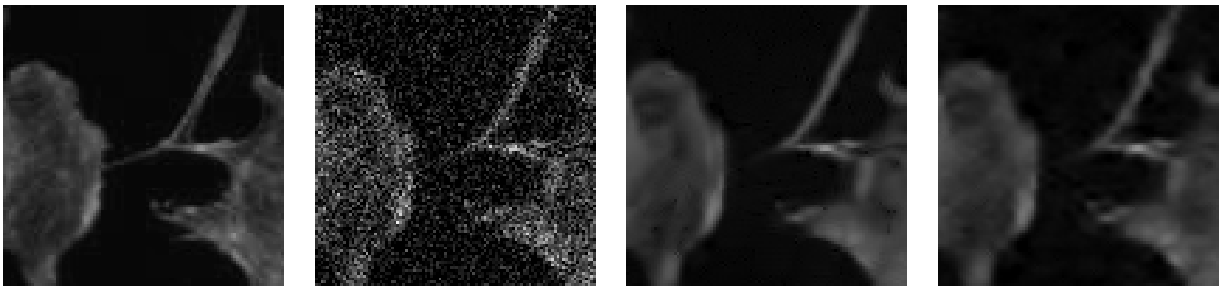
We denoise each image with the three-step variance stabilization approach explained in Section I, using either BM3D [10] or BLS-GSM [11] as the Gaussian denoising algorithm, and inverting the denoised data with each of the following transformations: the exact unbiased inverse $\mathcal{I}_\sigma$, its closed-form approximation $\widetilde{\mathcal{I}}_\sigma$, the asymptotically unbiased inverse $\mathcal{I}_{\text{asy}}(D) = \frac{1}{4}D^2 - \frac{1}{8} - \sigma^2$, or the algebraic inverse $\mathcal{I}_{\text{alg}}(D) = \frac{1}{4}D^2 - \frac{3}{8} - \sigma^2$. When denoising the stabilized data $f_\sigma(z)$, the algorithms assume that $\text{std}\,\{f_\sigma(z) \mid y,\sigma\}$ is exactly 1. For comparison with the direct approaches, we also denoise each image with the state-of-the-art UWT/BDCT PURE-LET method proposed in [4]. The results are presented in Tables I–II, where each PSNR value (Table I), and the respective SSIM [12] value (Table II), is an average of ten individual denoising results (performed on ten random realizations of the Poisson-Gaussian noise).

We see that the proposed method is competitive with the UWT/BDCT PURE-LET algorithm, outperforming it in many cases, in particular when variance stabilization is combined with the BM3D algorithm, which represents the state of the art in additive white Gaussian noise removal. Moreover, there are no major declines in performance for the low-intensity cases, which demonstrates the fact that the poor performance shown in earlier works (e.g., [4]) is not simply due to inadequate variance stabilization associated with the GAT, but mostly a consequence of using an improper inverse transformation. In particular, we see that for the low-intensity cases it is clearly not reasonable to use either the asymptotically unbiased inverse or the algebraic inverse; instead, the proposed exact unbiased inverse $\mathcal{I}_\sigma$ can be used everywhere. Further, its closed-form approximation $\widetilde{\mathcal{I}}_\sigma$ is practically on par with it, introducing minor discrepancies only at the lowest intensities.



(a) (b) (c) (d)

Fig. 5. The denoising of Fluorescent Cells (512×512). (a) Original image, (b) Noisy image (peak = 20, $\sigma = 2$, PSNR = 17.21 dB), (c) Denoised with BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$ (PSNR = 29.65 dB), (d) Denoised with UWT/BDCT PURE-LET (PSNR = 29.47 dB).



(a) (b) (c) (d)

Fig. 6. A 100×100 section of each of the images in Figure 5. (a) Original image, (b) Noisy image (peak = 20, $\sigma = 2$), (c) Denoised with BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$, (d) Denoised with UWT/BDCT PURE-LET.

(a) (b) (c) (d)

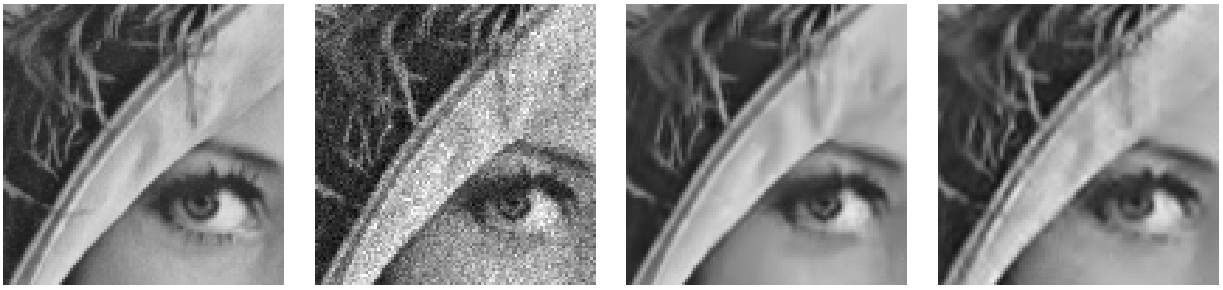
Fig. 7. The denoising of Lena (512×512). (a) Original image, (b) Noisy image (peak = 100, $\sigma = 2$, PSNR = 22.64 dB), (c) Denoised with BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$ (PSNR = 33.41 dB), (d) Denoised with UWT/BDCT PURE-LET (PSNR = 32.92 dB).

Figures 5–8 present visual comparisons for Fluorescent Cells (peak = 20, $\sigma = 2$) and Lena (peak = 100, $\sigma = 2$), corroborating the observed good performance of the proposed denoising method. Figure 9 shows the denoising results for the low-intensity case of Cameraman with peak = 1 and $\sigma = 0.1$, including a comparison of the different inverses. This clearly visualizes the previously noted importance of applying a proper inverse transformation to the denoised data.

B. Denoising with estimated parameter values α_{est} and $\overset{*}{\sigma}_{est}$

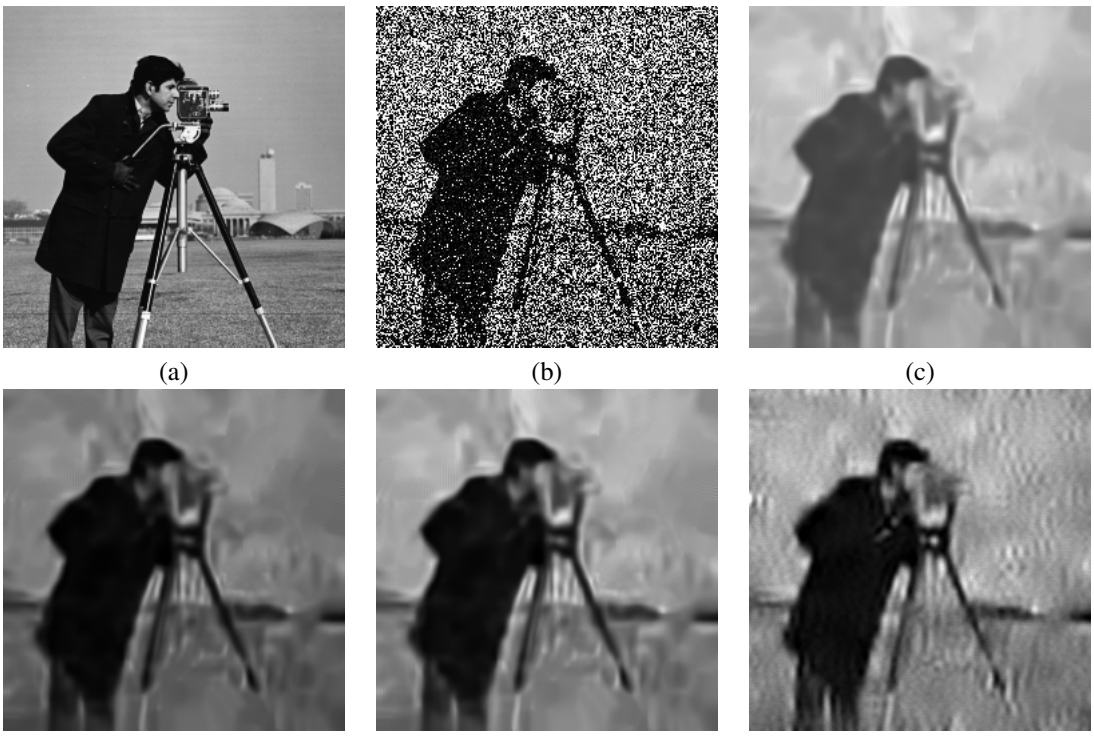
Here we examine the robustness and practical applicability of the proposed method by repeating a subset of the experiments corresponding to Tables I–II, but using estimated parameter values α_{est} and $\overset{*}{\sigma}_{\text{est}}$ instead of the true values α and $\overset{*}{\sigma}$. Note that since we now generally have $\alpha_{\text{est}} \neq 1$, we also have $\overset{*}{\sigma}_{\text{est}} \neq \sigma_{\text{est}}$. Thus, in this part of the experiments, we are making a distinction between $\overset{*}{\sigma}$ and σ for clarity, even though $\overset{*}{\sigma} = \sigma$ still holds, as in all the previous experiments. The parameters α_{est} and $\overset{*}{\sigma}_{\text{est}}$ are estimated from a single noisy image by fitting a global parametric model into locally estimated expectation / standard deviation pairs, as proposed in [1]. Then, the transformations (4) and the GAT (7) $f_{\sigma_{\text{est}}}$ are applied with $\sigma_{\text{est}} = \overset{*}{\sigma}_{\text{est}}/\alpha_{\text{est}}$, the stabilized data is denoised with BM3D, and the final estimate is obtained by inverting the denoised data with either the exact unbiased inverse $\mathcal{I}_{\sigma_{\text{est}}}$, the asymptotically unbiased inverse $\mathcal{I}_{\text{asy}}(D) = \frac{1}{4}D^2 - \frac{1}{8} - \sigma^2_{\text{est}}$, or the algebraic inverse $\mathcal{I}_{\text{alg}}(D) = \frac{1}{4}D^2 - \frac{3}{8} - \sigma^2_{\text{est}}$.

The results of these experiments are presented in Table III, showing the robustness of the proposed denoising framework, as in the vast majority of the cases the denoising results associated with $\mathcal{I}_{\sigma_{\text{est}}}$ are practically equal to the ones obtained with the exact parameter values α and σ (Tables I–II); the only notable declines in performance are the approximately 0.3 dB drops for Fluorescent Cells with peaks 1 and 2. However, these drops at low peaks are insignificant in comparison with the major decline (several dBs) in performance caused by using either the asymptotically unbiased inverse $\mathcal{I}_{\text{asy}}$ or the algebraic inverse $\mathcal{I}_{\text{alg}}$ instead of $\mathcal{I}_{\sigma_{\text{est}}}$.



(a) (b) (c) (d)

Fig. 8. A 100×100 section of each of the images in Figure 7. (a) Original image, (b) Noisy image (peak = 100, $\sigma = 2$), (c) Denoised with BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$, (d) Denoised with UWT/BDCT PURE-LET.



(a) (b) (c)

(d) (e) (f)

Fig. 9. The denoising of Cameraman (256×256). (a) Original image, (b) Noisy image (peak = 1, $\sigma = 0.1$, PSNR = 3.20 dB), (c) Denoised with BM3D and the asymptotically unbiased inverse $\mathcal{I}_{\text{asy}}$ (PSNR = 15.55 dB), (d) Denoised with BM3D and the algebraic inverse $\mathcal{I}_{\text{alg}}$ (PSNR = 15.72 dB), (e) Denoised with BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$ (PSNR = 20.23 dB), (f) Denoised with UWT/BDCT PURE-LET (PSNR = 20.35 dB).

We remark that each result in Table III is obtained by denoising a single image instead of averaging ten results, since in Table I we observed that the variations between individual results are typically minor (in the order of $\pm$ 0.1 dB around the average PSNR). Note that even though the true value of α equals 1 in all the experiments, this information is not used in any way in computing the estimates α_{est} and $\overset{*}{\sigma}_{\text{est}}$. Moreover, using only $\alpha = 1$ is not a fundamental restriction, as the more general case (1) can be addressed through simple scaling.

C. Varying the ratio between the Poisson and Gaussian noise components

Finally we analyze how much the denoising performance changes, when the noise distribution gradually changes from pure Poisson to additive white Gaussian noise (AWGN). In practice, we denoise Cameraman and Lena with either BM3D or BLS-GSM combined with the exact unbiased inverse $\mathcal{I}_\sigma$, and with UWT/BDCT PURE-LET, when the PSNR of the noisy image is kept constant at either 10 dB or 15 dB (with negligible variations depending on the specific realization, of the order $\pm$ 0.01 dB), but the ratio $\sigma/\sqrt{\text{peak}}$ is varied from 0 to 10. When this ratio is low, the Poisson noise component is the dominant one, whereas a high ratio means the noise distribution is nearly Gaussian. Specifically, the GAT corresponding to the case $\sigma/\sqrt{\text{peak}} = 10$ is practically affine over the interval [0, peak], and thus, this case is essentially like denoising pure AWGN without any variance stabilization. Conversely, when $\sigma/\sqrt{\text{peak}} = 0$, the Gaussian component is absent, and hence we reduce to the case of denoising pure Poisson data [7].

Figures 10–11 show the results of these experiments. First, we see the general trend that the denoising performance increases as the distribution becomes less Gaussian. We explain

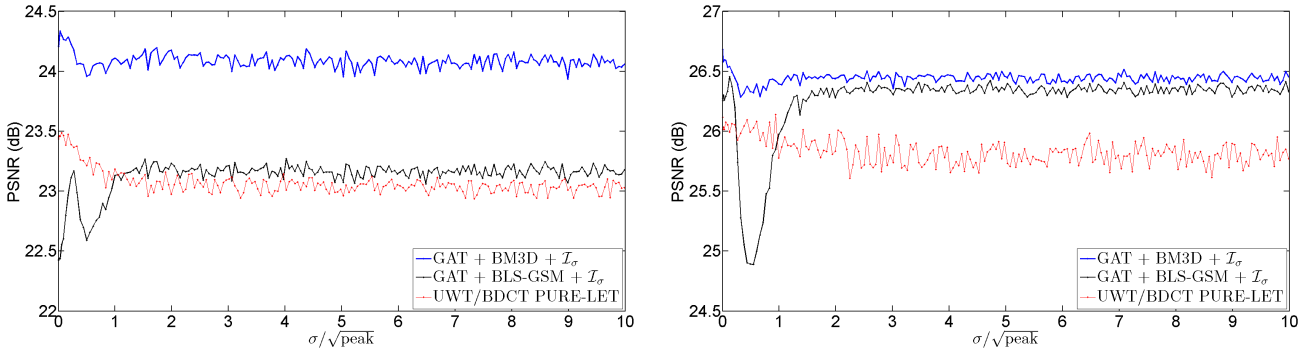


Fig. 10. The results of denoising variance-stabilized data with either BM3D or BLS-GSM combined with the exact unbiased inverse $\mathcal{I}_\sigma$, and denoising with UWT/BDCT PURE-LET, when the PSNR of the noisy image is kept constant at 10 dB, but the ratio $\sigma/\sqrt{\text{peak}}$ is varied from 0 (i.e., the Poisson noise component dominates) to 10 (i.e., the Gaussian noise component dominates). Each PSNR value is an average over five random noise realizations. (a) Cameraman (256×256); peak = 4.71 when $\sigma = 0$, (b) Lena (512×512); peak = 5.05 when $\sigma = 0$.

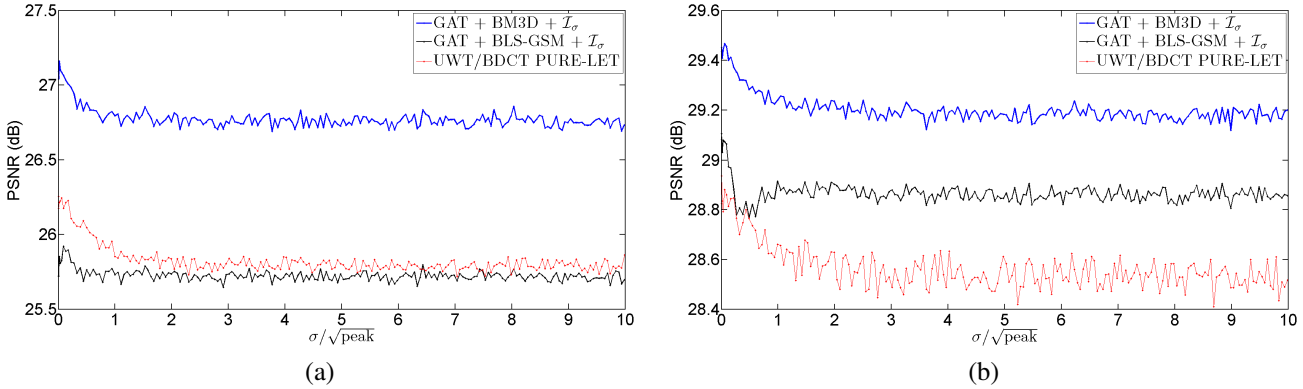


Fig. 11. The results of denoising variance-stabilized data with either BM3D or BLS-GSM combined with the exact unbiased inverse $\mathcal{I}_\sigma$, and denoising with UWT/BDCT PURE-LET, when the PSNR of the noisy image is kept constant at 15 dB, but the ratio $\sigma/\sqrt{\text{peak}}$ is varied from 0 (i.e., the Poisson noise component dominates) to 10 (i.e., the Gaussian noise component dominates). Each PSNR value is an average over five random noise realizations. (a) Cameraman (256×256); peak = 14.90 when $\sigma = 0$, (b) Lena (512×512); peak = 15.95 when $\sigma = 0$.

this perhaps curious behaviour by the fact that a Gaussian distribution can be considered the worst-case scenario, in the sense that it leads to the largest Cramér-Rao lower bound under a rather general class of distributions [13]. Second, it is apparent from Figure 10(b) that in the 10 dB case, there is a minor drop in performance for BM3D at about $\sigma/\sqrt{\text{peak}} = 0.5$, but it is not observed for UWT/BDCT PURE-LET. This deficiency could be caused by inaccurate variance stabilization. In particular, as $\sigma/\sqrt{\text{peak}}$ decreases, the stabilization changes from nearly perfect to a slight overshooting (up to about 10 % extra in terms of the standard deviation), and finally undershooting (see Figure 1(b)). For low input PSNR values, such as 10 dB, the intensity range of the scaled image is typically concentrated in the region with inaccurate stabilization. However, for higher PSNR values this concerns only the darkest part of the image. This notion is supported by the comparisons in Figure 11, showing the corresponding results for the 15 dB case; there is no drop in performance for BM3D due to variance stabilization. Overall, BM3D and the exact unbiased inverse $\mathcal{I}_\sigma$ outperforms the state-of-the-art UWT/BDCT PURE-LET in all cases, even when inaccurate variance stabilization causes a minor drop in performance for low input PSNR values around a specific $\sigma/\sqrt{\text{peak}}$ ratio.

The performance of BLS-GSM is much degraded by the mismatch between the actual standard deviation of the stabilized data and its assumed value (which is always 1). This is not a peculiarity of the Poisson-Gaussian case, as drops in performance comparable to those visible in Figure 10 can be experienced with BLS-GSM also in the case of purely Gaussian noise when the standard deviation of the noise is over- or underestimated.

D. Computational complexity

To give the reader an indication of the typical execution times, for Lena (512×512) the various components of the denoising process require time approximately as follows: GAT 0.005 s, BM3D 5.4 s, exact unbiased inverse $\mathcal{I}_\sigma$ 0.2 s, closed-form approximation $\widetilde{\mathcal{I}}_\sigma$ 0.05 s, and the asymptotically unbiased inverse $\mathcal{I}_{\text{asy}}$ 0.005 s. Hence, for instance, denoising Lena with GAT + BM3D + exact unbiased inverse $\mathcal{I}_\sigma$ takes about 5.6 s in total. Note that all of these execution times scale linearly with respect to the image size. If BLS-GSM is used instead of BM3D, one can expect a tenfold increase in the denoising time. Finally, denoising Lena with UWT/BDCT PURE-LET requires approximately 42 s (with 16×16 DCT blocks, as used in all of our experiments for the best denoising results). These results are obtained with an Intel Core 2 Duo E8400 processor, running at 3.0 GHz.

TABLE I
A COMPARISON OF THE DENOISING PERFORMANCE (PSNR, DB) OF SEVERAL DENOISING ALGORITHMS AND INVERSE TRANSFORMATIONS, USING VARIOUS PEAK INTENSITIES AND VARIOUS STANDARD DEVIATIONS σ OF THE GAUSSIAN NOISE COMPONENT. THE RESULTS ARE AVERAGES OF TEN INDEPENDENT REALIZATIONS, AND THE VARIATIONS BETWEEN INDIVIDUAL RESULTS ARE TYPICALLY IN THE ORDER OF $\pm$ 0.1 DB AROUND THE AVERAGE PSNR.

Image	Peak	σ	Noisy	GAT + BM3D				GAT + BLS-GSM				UWT/BDCT
				$\mathcal{I}_\sigma$	$\widetilde{\mathcal{I}}_\sigma$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$	$\mathcal{I}_\sigma$	$\widetilde{\mathcal{I}}_\sigma$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$	PURE-LET [4]
Cameraman (256×256)	1	0.1	3.20	20.23	20.18	15.55	15.72	18.46	18.45	14.56	15.40	**20.35**
	2	0.2	6.12	**21.93**	21.85	20.70	18.24	20.28	20.23	19.41	17.36	21.60
	5	0.5	9.83	**24.09**	24.07	24.00	22.36	23.01	22.98	22.93	21.29	23.33
	10	1	12.45	**25.52**	**25.52**	**25.52**	24.80	24.36	24.36	24.35	23.62	24.68
	20	2	14.76	**26.77**	**26.77**	26.75	26.48	25.58	25.58	25.53	25.26	25.92
	30	3	15.91	**27.30**	**27.30**	27.29	27.13	26.20	26.20	26.16	26.01	26.51
	60	6	17.49	**28.07**	**28.07**	28.06	28.01	27.02	27.02	26.98	26.93	27.35
	120	12	18.57	**28.57**	**28.57**	28.55	28.54	27.57	27.57	27.52	27.51	27.89
Fluorescent Cells (512×512)	1	0.1	7.22	24.54	24.45	13.86	20.83	22.35	22.33	13.47	20.01	**25.13**
	2	0.2	9.99	25.87	25.60	20.99	21.96	24.20	24.07	20.25	21.23	**26.25**
	5	0.5	13.37	27.45	27.38	26.93	24.80	26.99	26.88	26.52	24.25	**27.60**
	10	1	15.53	28.63	**28.64**	28.61	27.20	28.05	28.04	28.03	26.50	28.59
	20	2	17.21	29.65	**29.66**	29.64	29.09	29.05	29.06	28.89	28.28	29.47
	30	3	17.97	**30.16**	**30.16**	30.15	29.86	29.74	29.74	29.65	29.33	29.84
	60	6	18.86	**30.77**	**30.77**	**30.77**	30.68	30.52	30.52	30.48	30.38	30.42
	120	12	19.39	**31.14**	**31.14**	**31.14**	31.11	30.91	30.91	30.87	30.85	30.70
Lena (512×512)	1	0.1	2.87	22.59	22.50	16.89	16.38	21.55	21.54	16.04	16.48	**22.83**
	2	0.2	5.82	**24.34**	24.21	23.31	18.91	23.56	23.46	22.52	18.58	24.16
	5	0.5	9.54	26.17	26.16	**26.18**	23.52	25.98	25.94	25.98	22.91	25.74
	10	1	12.19	27.72	27.71	**27.73**	26.50	26.85	26.84	26.84	25.57	27.27
	20	2	14.53	**29.01**	29.00	**29.01**	28.54	28.47	28.47	28.47	27.96	28.46
	30	3	15.72	**29.69**	**29.69**	**29.69**	29.44	29.26	29.27	29.26	28.99	29.12
	60	6	17.35	**30.51**	**30.51**	**30.51**	30.43	30.12	30.12	30.12	30.03	29.91
	120	12	18.48	**31.05**	**31.05**	**31.05**	31.03	30.69	30.69	30.69	30.66	30.51

Based on the above results, and on the notion that the proposed exact unbiased inverse $\mathcal{I}_\sigma$ is conceptually similar to the exact unbiased inverse of the Anscombe transformation introduced in our earlier work [7], it is evident that the computational complexity associated with the exact unbiased inverse is similar in both scenarios. Thus, we will not delve deeper on this issue, and refer the reader to [7, Table V] for complementary benchmarks. The important thing to emphasize is that the time required for the whole denoising process is overwhelmingly determined by the execution time of the chosen Gaussian denoising algorithm. The components of the variance stabilization process (applying the affine transformations (4) and the corresponding affine inverse, the GAT, and the exact unbiased inverse) are very simple operations in comparison, keeping in mind that the exact unbiased inverse is implemented either via a pre-computed lookup table or via the closed-form approximation (21).

V. DISCUSSION AND CONCLUSIONS

We have generalized our earlier work [7], [8] in order to encompass the case of Poisson-Gaussian noise. Specifically, we proposed an exact unbiased inverse of the generalized Anscombe transformation for Poisson-Gaussian noise and supplemented it with rigorous mathematical considerations. We also proposed a closed-form approximation of this inverse, based on our previously presented approximation for the exact unbiased inverse of the Anscombe transformation [8].

We showed that the denoising performance associated with the proposed exact unbiased inverse, in conjunction with a state-of-the-art Gaussian noise removal algorithm, is competitive with that of a state-of-the-art algorithm designed specifically for the removal of Poisson-Gaussian noise. Further, we observed that for low peak intensities, the performance gain obtained by using the exact unbiased inverse instead of the algebraic or the asymptotically unbiased inverse is especially significant. In other words, we showed that the poor denoising performance shown in earlier works is not simply

TABLE II
A COMPARISON OF THE DENOISING PERFORMANCE (MEAN SSIM [12]) OF SEVERAL DENOISING ALGORITHMS AND INVERSE TRANSFORMATIONS, USING VARIOUS PEAK INTENSITIES AND VARIOUS STANDARD DEVIATIONS σ OF THE GAUSSIAN NOISE COMPONENT. THE RESULTS ARE AVERAGES OF TEN INDEPENDENT REALIZATIONS (EXACTLY THE SAME REALIZATIONS AS IN TABLE I).

Image	Peak	σ	Noisy	GAT + BM3D				GAT + BLS-GSM				UWT/BDCT
				$\mathcal{I}_\sigma$	$\widetilde{\mathcal{I}}_\sigma$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$	$\mathcal{I}_\sigma$	$\widetilde{\mathcal{I}}_\sigma$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$	PURE-LET [4]
Cameraman (256×256)	1	0.1	0.0442	**0.5991**	0.5973	0.5071	0.5652	0.5204	0.5177	0.4579	0.5007	0.5232
	2	0.2	0.0736	**0.6512**	0.6427	0.5979	0.5992	0.5873	0.5784	0.5470	0.5437	0.5573
	5	0.5	0.1238	**0.7068**	0.7044	0.6998	0.6663	0.6569	0.6528	0.6551	0.6131	0.6227
	10	1	0.1661	0.7509	**0.7514**	0.7512	0.7343	0.6689	0.6685	0.6697	0.6419	0.6849
	20	2	0.2085	0.7882	**0.7883**	0.7879	0.7840	0.7135	0.7131	0.7098	0.7012	0.7339
	30	3	0.2314	**0.8024**	**0.8024**	0.8022	0.8005	0.7414	0.7413	0.7392	0.7348	0.7554
	60	6	0.2657	**0.8225**	**0.8225**	0.8224	0.8218	0.7696	0.7696	0.7682	0.7666	0.7832
	120	12	0.2913	0.8345	**0.8346**	0.8344	0.8342	0.7865	0.7866	0.7851	0.7845	0.7992
Fluorescent Cells (512×512)	1	0.1	0.0234	0.6090	0.5938	0.3465	0.5189	0.5098	0.5002	0.2939	0.4715	**0.6230**
	2	0.2	0.0421	0.6558	0.6031	0.4828	0.4518	0.5759	0.5302	0.4149	0.4169	**0.6567**
	5	0.5	0.0806	**0.7120**	0.6943	0.6634	0.4930	0.6821	0.6558	0.6379	0.4421	0.6982
	10	1	0.1184	0.7491	**0.7515**	0.7464	0.6134	0.7018	0.6993	0.7018	0.5148	0.7288
	20	2	0.1569	0.7795	**0.7811**	0.7788	0.7365	0.7301	0.7292	0.7093	0.6347	0.7566
	30	3	0.178	**0.7980**	**0.7980**	0.7968	0.7750	0.7608	0.7605	0.7488	0.7098	0.7661
	60	6	0.2073	0.8157	**0.8158**	0.8152	0.8082	0.7913	0.7915	0.7861	0.7728	0.7867
	120	12	0.2267	**0.8241**	**0.8241**	0.8236	0.8207	0.8021	0.8022	0.7973	0.7919	0.7913
Lena (512×512)	1	0.1	0.016	0.6288	**0.6290**	0.6108	0.6068	0.6134	0.6135	0.5899	0.5929	0.5767
	2	0.2	0.0296	**0.6676**	0.6670	0.6657	0.6513	0.6606	0.6604	0.6561	0.6443	0.6031
	5	0.5	0.0606	**0.7038**	0.7031	**0.7038**	0.6985	0.6812	0.6804	0.6813	0.6754	0.6406
	10	1	0.0946	**0.7554**	0.7553	0.7553	0.7538	0.6860	0.6855	0.6859	0.6841	0.7227
	20	2	0.1343	**0.7934**	**0.7934**	**0.7934**	0.7930	0.7622	0.7621	0.7621	0.7616	0.7662
	30	3	0.1591	**0.8107**	**0.8107**	**0.8107**	0.8105	0.7915	0.7915	0.7915	0.7913	0.7882
	60	6	0.1965	**0.8297**	**0.8297**	**0.8297**	**0.8297**	0.8159	0.8159	0.8159	0.8158	0.8088
	120	12	0.2266	**0.8410**	**0.8410**	**0.8410**	**0.8410**	0.8298	0.8298	0.8298	0.8298	0.8264

due to the inability of the GAT to stabilize the noise variance adequately, but mostly due to applying an unsuitable inverse transformation.

We also wish to point out that the GAT is not the only possible choice for stabilizing the noise variance. For instance, one may instead compute a family of optimized transformations [14] (mitigating the undershoot and overshoot observed in Figure 1(b)), to which our denoising framework can be accommodated by recomputing the grid of corresponding expected values (10). However, this is not within the scope of this paper, which focuses on the major significance of the exact unbiased inverse transformation, not on optimized variance stabilization.

In light of our contributions, it seems questionable whether it is advantageous to go through the effort of designing separate denoising algorithms for each specific noise distribution, when comparable gains can be attained by perfecting the denoising algorithms in the AWGN case and addressing various noise distributions through variance stabilization together with a suitable inverse.

VI. ACKNOWLEGDEMENT

We thank Florian Luisier for providing us with their PURE-LET software, along with the Fluorescent Cells test image used in the experiments. We also thank the reviewers for their feedback that helped us in improving this paper.

REFERENCES

[1] Foi, A., M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw data", *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008. doi:10.1109/TIP.2008.2001399

[2] Foi, A., "Clipped noisy images: heteroskedastic modeling and practical denoising", *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009. doi:10.1016/j.sigpro.2009.04.035

[3] Zhang, B., M.J. Fadili, J.L. Starck, and J.C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging", *Proc. IEEE Int. Conf. Image Process., ICIP 2007*, pp. VI-234–VI-236, 2007.

[4] Luisier, F., T. Blu, and M. Unser, "Image Denoising in Mixed Poisson-Gaussian Noise", *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, March 2011.

[5] Mäkitalo, M., and A. Foi, "Poisson-Gaussian denoising using the exact unbiased inverse of the generalized Anscombe transformation", *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP 2012*, March 2012.

TABLE III
THE DENOISING PERFORMANCE (PSNR (DB), AND MEAN SSIM [12]) OF BM3D, COMBINED WITH EITHER THE EXACT UNBIASED INVERSE TRANSFORMATION $\mathcal{I}_\sigma$, THE ASYMPTOTICALLY UNBIASED INVERSE $\mathcal{I}_{\text{ASY}}$ OR THE ALGEBRAIC INVERSE $\mathcal{I}_{\text{ALG}}$, USING ESTIMATED PARAMETER VALUES α_{EST} AND $\overset{*}{\sigma}_{\text{EST}}$ INSTEAD OF THE TRUE VALUES $\alpha = 1$ AND $\overset{*}{\sigma}$. EACH RESULT IS OBTAINED FROM A SINGLE NOISY REALIZATION, BUT AS THE VARIATIONS BETWEEN INDIVIDUAL RESULTS ARE TYPICALLY MINOR, THE RESULTS ARE COMPARABLE WITH THE AVERAGE RESULTS PRESENTED IN TABLES I–II. NOTE THAT SINCE $\alpha_{\text{EST}} \neq 1$, WE HAVE $\overset{*}{\sigma}_{\text{EST}} \neq \sigma_{\text{EST}}$, EVEN THOUGH $\overset{*}{\sigma} = \sigma$.

Image	Peak	$\overset{*}{\sigma} = \sigma$	α_{est}	$\overset{*}{\sigma}_{\text{est}}$	$\sigma_{\text{est}} = \frac{\overset{*}{\sigma}_{\text{est}}}{\alpha_{\text{est}}}$	PSNR			SSIM		
						$\mathcal{I}_{\sigma_{\text{est}}}$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$	$\mathcal{I}_{\sigma_{\text{est}}}$	$\mathcal{I}_{\text{asy}}$	$\mathcal{I}_{\text{alg}}$
Cameraman (256×256)	1	0.1	0.957	0.141	0.147	**20.34**	16.04	15.71	**0.5923**	0.5039	0.5614
	2	0.2	0.948	0.118	0.125	**21.77**	21.00	18.23	**0.5994**	0.5691	0.5544
	5	0.5	1.015	0.459	0.453	**24.19**	24.10	22.38	**0.7115**	0.7039	0.6705
	10	1	1.095	0.896	0.819	**25.51**	**25.51**	24.61	0.7600	**0.7604**	0.7393
	20	2	1.008	2.075	2.059	**26.81**	26.79	26.54	**0.7888**	0.7887	0.7857
	30	3	1.053	3.038	2.885	**27.24**	27.23	27.09	**0.8048**	0.8045	0.8028
	60	6	0.973	6.224	6.398	**28.08**	28.07	28.01	**0.8203**	0.8201	0.8195
	120	12	0.907	12.505	13.781	**28.64**	28.62	28.60	**0.8379**	0.8378	0.8371
Fluorescent Cells (512×512)	1	0.1	0.949	0.183	0.193	**24.19**	14.40	20.81	**0.5655**	0.3541	0.5191
	2	0.2	0.920	0.153	0.166	**25.55**	22.07	21.58	**0.6184**	0.5001	0.4184
	5	0.5	0.984	0.483	0.491	**27.54**	27.06	24.78	**0.7140**	0.6678	0.4947
	10	1	0.993	0.995	1.002	**28.62**	28.60	27.18	**0.7503**	0.7476	0.6211
	20	2	0.960	2.035	2.121	**29.71**	29.70	29.19	**0.7827**	0.7813	0.7378
	30	3	0.928	3.067	3.305	**30.26**	**30.26**	30.01	**0.8040**	0.8036	0.7879
	60	6	0.994	5.993	6.027	**30.79**	30.78	30.70	**0.8182**	0.8179	0.8122
	120	12	1.143	12.014	10.516	**31.16**	**31.16**	31.13	**0.8244**	0.8239	0.8205
Lena (512×512)	1	0.1	1.003	0.218	0.218	**22.60**	16.62	16.62	**0.6303**	0.6107	0.6109
	2	0.2	1.019	0.200	0.196	**24.35**	23.17	18.93	**0.6697**	0.6672	0.6535
	5	0.5	1.074	0.285	0.265	26.09	**26.12**	23.31	0.7054	**0.7059**	0.6995
	10	1	0.992	1.086	1.094	**27.78**	27.77	26.61	**0.7571**	0.7570	0.7556
	20	2	1.091	1.884	1.726	**29.06**	**29.06**	28.56	**0.7976**	**0.7976**	0.7971
	30	3	0.994	3.100	3.118	**29.81**	**29.81**	29.61	**0.8136**	**0.8136**	0.8134
	60	6	1.068	5.861	5.489	**30.55**	**30.55**	30.47	**0.8308**	**0.8308**	0.8307
	120	12	0.792	12.490	15.778	**31.02**	**31.02**	31.01	**0.8393**	**0.8393**	**0.8393**

[6] Starck, J.L., F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis*, Cambridge University Press, Cambridge, 1998.

[7] Mäkitalo, M., and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising", *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, Jan. 2011. http://dx.doi.org/10.1109/TIP.2010.2056693

[8] Mäkitalo, M., and A. Foi, "A closed-form approximation of the exact unbiased inverse of the Anscombe variance-stabilizing transformation", *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2697–2698, Sep. 2011. http://dx.doi.org/10.1109/TIP.2011.2121085

[9] Anscombe, F.J., "The transformation of Poisson, binomial and negative-binomial data", *Biometrika*, vol. 35, no. 3/4, pp. 246–254, Dec. 1948. http://dx.doi.org/10.1093/biomet/35.3-4.246

[10] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering", *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[11] Portilla, J., V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain", *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[12] Wang, Z., A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

[13] Stoica, P., and P. Babu, "The Gaussian Data Assumption Leads to the Largest Cramér-Rao Bound", *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 132–133, May 2011.

[14] Foi, A., "Optimization of variance-stabilizing transformations", 2009, preprint available at http://www.cs.tut.fi/~foi/.

[15] Kendall, M., *The advanced theory of statistics*, *Volume 1*, 2nd ed., Charles Griffin and co., 1945.

APPENDIX

In Section A, we analyze the asymptotic behaviour of the exact unbiased inverse $\mathcal{I}_\sigma$ of the generalized Anscombe transformation f_σ (7) for large and small values of σ, as well as for large values of its argument D. Further, as a result of our analysis, we also obtain a closed-form approximation of $\mathcal{I}_\sigma$, which can be used for any value of D and σ.

Finally, in Section B we prove (14), showing that the exact unbiased inverse $\mathcal{I}_\sigma$ coincides with a maximum likelihood inverse.

A. Asymptotics for the exact unbiased inverse

1) Expansion about the mean: Let

$$f_\sigma(z) = \begin{cases} 2\sqrt{z + \frac{3}{8} + \sigma^2}, & z > -\frac{3}{8} - \sigma^2 \\ 0, & z \leq -\frac{3}{8} - \sigma^2 \end{cases} . \qquad (25)$$

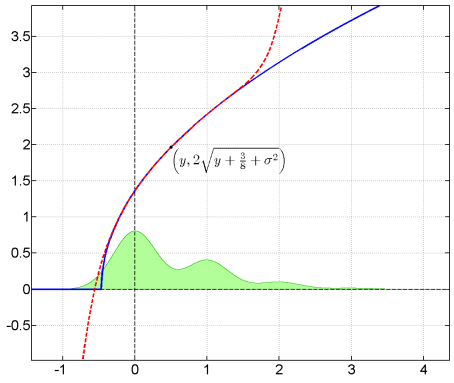


Fig. 12. Probability density function of z when $y = 0.5$ and $\sigma = 0.3$ (green area plot); generalized Anscombe transformation f_σ (blue solid line) and its Taylor expansion of order 9 centered at y (red dashed line).

Following [9], we consider the Taylor expansion of $f_\sigma(z)$ about the mean $y = E\{z\} \geq 0$ of z. Defining $t = z - y$ and $\breve{y} = y + \frac{3}{8} + \sigma^2$, we have

$$f_\sigma(z) = \begin{cases} 2\sqrt{\breve{y}}\left(1 + \frac{1}{2}\frac{t}{\breve{y}} - \frac{1}{8}\left(\frac{t}{\breve{y}}\right)^2 + \right. & \\ \quad + \frac{1}{16}\left(\frac{t}{\breve{y}}\right)^3 - \frac{5}{128}\left(\frac{t}{\breve{y}}\right)^4 + \ldots & t > -\breve{y}, \\ \quad \left. \ldots + (-1)^s a_{s-1}\left(\frac{t}{\breve{y}}\right)^{s-1} + R_s\right) & \\ 0 & t \leq -\breve{y}, \end{cases} \tag{26}$$

where

$$a_s = (-1)^{s+1}\frac{1\cdot(-1)\cdot(-3)\cdot\ldots\cdot(-2s+3)}{2^s s!} = \\ = -\prod_{j=1}^{s} 1 - \frac{3}{2j} = \mathcal{O}\left(\frac{1}{s^{3/2}}\right)$$

and the series converges for $|t| \leq \breve{y}$, with the remainder term R_s being

$$R_s = \mathcal{O}\left(\left(\frac{t}{\breve{y}}\right)^s\right)$$

for all $t \in \mathbb{R}$. This expansion is illustrated in Figure 12.

2) Moments: Let μ_k be the k-th centered moment of z. In particular,

$$\mu_1 = 0, \quad \mu_2 = y + \sigma^2, \quad \mu_3 = y, \quad \mu_4 = y + 3\left(y + \sigma^2\right)^2.$$

More in general, μ_k can be expressed as a polynomial of order at most $\frac{k}{2}$ with non-negative coefficients in the $k-1$ cumulants $\{\kappa_j\}_{j=2}^k$ of z (see [15], Sect. 3.13, pp. 61–63). In turn, each cumulant κ_j can expressed as the sum of the j-th cumulants $\kappa_j^{\mathcal{P}(y)}$ and $\kappa_j^{\mathcal{N}\left(0,\sigma^2\right)}$ of the Poisson $\mathcal{P}(y)$ and Gaussian distribution $\mathcal{N}\left(0,\sigma^2\right)$, respectively. In particular, for the Poisson we have $\kappa_j^{\mathcal{P}(y)} = y\ \forall j$, while for the Gaussian $\kappa_2^{\mathcal{N}\left(0,\sigma^2\right)} = \sigma^2$ and $\kappa_j^{\mathcal{N}\left(0,\sigma^2\right)} = 0\ \forall j \neq 2$ [15]. This means that $\kappa_2 = y + \sigma^2$ and $\kappa_j = y\ \forall j \neq 2$, and hence that μ_k is a polynomial in y and $y + \sigma^2$, which we denote as

$$\mu_k = P_k\left(y, y + \sigma^2\right). \tag{27}$$

Let us consider now the order of the centered moments and absolute centered moments. As observed in [9], the k-th absolute centered moment $\nu_k^{\mathcal{P}(m)}$ of a Poisson variable with mean m is of order $\mathcal{O}\left(m^{\frac{1}{2}k}\right)$ as $m \to +\infty$. The same order applies also to the centered moments $\mu_k^{\mathcal{P}(m)}$ of the same Poisson variable. Since $\kappa_j \leq y + \sigma^2 < y + \frac{3}{8} + \sigma^2 = \breve{y}$ and because moments are polynomials with non-negative coefficients in the cumulants, we have that

$$\mu_k^{\mathcal{P}(y)} \leq \mu_k \leq \mu_k^{\mathcal{P}\left(y+\sigma^2\right)} \leq \mu_k^{\mathcal{P}(\breve{y})},$$

whence we obtain that μ_k is of at most order $\mathcal{O}\left(\breve{y}^{\frac{1}{2}k}\right)$ as $\breve{y} \to +\infty$. The same asymptotic order extends to the absolute moments ν_k, because, for any odd k, $\nu_{k-1} = \mu_{k-1}$ and $\nu_k \leq \left(\mu_{k+1}\right)^{k/(k+1)}$ (see [15], Sect. 3.6, p. 56).

3) Inverse mappings: Taking expectations on both sides of (26), we obtain

$$E\{f_\sigma(z)\,|y,\sigma\} = 2\sqrt{\breve{y}}\left(1 - \frac{1}{8}\frac{\mu_2}{\breve{y}^2} + \frac{1}{16}\frac{\mu_3}{\breve{y}^3} - \frac{5}{128}\frac{\mu_4}{\breve{y}^4} + \ldots \right. \\ \left. \ldots + (-1)^s a_{s-1}\frac{\mu_{s-1}}{\breve{y}^{s-1}} + \mathcal{O}\left(\frac{\mu_s}{\breve{y}^s}\right)\right). \tag{28}$$

By its very definition, the exact unbiased inverse $\mathcal{I}_\sigma$ of the generalized Anscombe transformation f_σ (7) is the mapping

$$E\{f_\sigma(z)\,|y,\sigma\} \overset{\mathcal{I}_\sigma}{\longmapsto} y,$$

i.e.,

$$2\sqrt{y+\tfrac{3}{8}+\sigma^2}\left(1 - \frac{1}{8}\frac{y+\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^2} + \frac{1}{16}\frac{y}{\left(y+\frac{3}{8}+\sigma^2\right)^3} + \right. \\ - \frac{5}{128}\frac{y+3\left(y+\sigma^2\right)^2}{\left(y+\frac{3}{8}+\sigma^2\right)^4} + \ldots + (-1)^s a_{s-1}\frac{P_{s-1}\left(y, y+\sigma^2\right)}{\left(y+\frac{3}{8}+\sigma^2\right)^{s-1}} + \\ \left. + \mathcal{O}\left(\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{s/2}}\right)\right) \overset{\mathcal{I}_\sigma}{\longmapsto} y. \tag{29}$$

For the particular case $\sigma = 0$, the above inverse reduces to the exact unbiased inverse $\mathcal{I}_0$ [7] of the Anscombe transformation f_0, which is hence the mapping

$$2\sqrt{y+\tfrac{3}{8}}\left(1 - \frac{1}{8}\frac{y}{\left(y+\frac{3}{8}\right)^2} + \frac{1}{16}\frac{y}{\left(y+\frac{3}{8}\right)^3} + \right. \\ - \frac{5}{128}\frac{y+3y^2}{\left(y+\frac{3}{8}\right)^4} + \ldots + (-1)^s a_{s-1}\frac{P_{s-1}\left(y, y\right)}{\left(y+\frac{3}{8}\right)^{s-1}} + \\ \left. + \mathcal{O}\left(\frac{1}{\left(y+\frac{3}{8}\right)^{s/2}}\right)\right) \overset{\mathcal{I}_0}{\longmapsto} y. \tag{30}$$

If we substitute y with $y + \sigma^2$ in (30), we obtain

$$2\sqrt{y+\tfrac{3}{8}+\sigma^2}\left(1 - \frac{1}{8}\frac{y+\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^2} + \frac{1}{16}\frac{y+\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^3} + \right. \\ - \frac{5}{128}\frac{y+\sigma^2+3\left(y+\sigma^2\right)^2}{\left(y+\frac{3}{8}+\sigma^2\right)^4} + \ldots + (-1)^s a_{s-1}\frac{P_{s-1}\left(y+\sigma^2, y+\sigma^2\right)}{\left(y+\frac{3}{8}+\sigma^2\right)^{s-1}} + \\ \left. + \mathcal{O}\left(\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{s/2}}\right)\right) \overset{\mathcal{I}_0}{\longmapsto} y+\sigma^2. \tag{31}$$

Of course, (30) and (31) define the same inverse $\mathcal{I}_0$, but the similarities between $\mathcal{I}_\sigma$ and $\mathcal{I}_0$ become more evident when comparing the mapping (29) with (31). Indeed, if we subtract σ^2 from $\mathcal{I}_0$, we obtain a mapping which coincides with $\mathcal{I}_\sigma$ except for some high-order terms in the argument of the

mapping:

$$2\sqrt{y+\tfrac{3}{8}+\sigma^2}\left(1-\frac{1}{8}\frac{y+\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^2}+\frac{1}{16}\frac{y+\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^3}+\right.$$
$$-\frac{5}{128}\frac{y+\sigma^2+3\left(y+\sigma^2\right)^2}{\left(y+\frac{3}{8}+\sigma^2\right)^4}+\ldots+(-1)^s a_{s-1}\frac{P_{s-1}\left(y+\sigma^2,y+\sigma^2\right)}{\left(y+\frac{3}{8}+\sigma^2\right)^{s-1}}+$$
$$\left.+\mathcal{O}\left(\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{s/2}}\right)\right)\overset{\mathcal{I}_0-\sigma^2}{\longmapsto} y. \quad (32)$$

The mappings (29) and (32) can be written in more compact form as

$$\mathcal{I}_\sigma\left(E\left\{f_\sigma\left(z\right)|y,\sigma\right\}\right)=y, \quad (33)$$
$$\mathcal{I}_0\left(E\left\{f_\sigma\left(z\right)|y,\sigma\right\}+\delta\left(y,\sigma\right)\right)-\sigma^2=y, \quad (34)$$

where $\delta\left(y,\sigma\right)$ is the difference between the left-hand sides of (32) and (29):

$$\delta\left(y,\sigma\right)=\frac{1}{8}\frac{\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^{5/2}}-\frac{5}{64}\frac{\sigma^2}{\left(y+\frac{3}{8}+\sigma^2\right)^{7/2}}+\ldots$$
$$\ldots+(-1)^s a_{s-1}\cdot 2\cdot\frac{P_{s-1}\left(y+\sigma^2,y+\sigma^2\right)-P_{s-1}\left(y,y+\sigma^2\right)}{\left(y+\frac{3}{8}+\sigma^2\right)^{s-3/2}}+$$
$$+\mathcal{O}\left(\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{(s-1)/2}}\right). \quad (35)$$

From (35) we can immediately see that $\delta\left(y,\sigma\right)\to 0$ as $y+\frac{3}{8}+\sigma^2\to+\infty$. In particular, $\delta\left(y,\sigma\right)=\mathcal{O}\left(\sigma^{-3}\right)$ as $\sigma\to+\infty$ and $\delta\left(y,\sigma\right)=\mathcal{O}\left(y^{-5/2}\right)$ as $y\to+\infty$ for a fixed σ.

To consider the case of small σ, let us first show that we can factor out σ^2 from the difference $P_k\left(y+\sigma^2,y+\sigma^2\right)-P_k\left(y,y+\sigma^2\right)$. It suffices to examine the differences $d_{i,j}$ between the corresponding monomials of equal order,

$$d_{i,j}=\left(y+\sigma^2\right)^{i+j}-y^i\left(y+\sigma^2\right)^j \qquad i+j\leq k-1$$

and observe that, upon expanding the powers, both the left term (minuend) and the right term (subtrahend) yield only one monomial which does not have a factor σ^2, namely y^{i+j}, and that this monomial vanishes in the subtraction. Thus,

$$P_k\left(y+\sigma^2,y+\sigma^2\right)-P_k\left(y,y+\sigma^2\right)=\sigma^2 Q_k\left(y,\sigma^2\right),$$

where $Q_k\left(y,\sigma^2\right)$ is polynomial in y and σ^2, again of the order $\mathcal{O}\left(\left(y+\frac{3}{8}+\sigma^2\right)^{\frac{1}{2}k}\right)$ as $y+\frac{3}{8}+\sigma^2\to+\infty$ and $\mathcal{O}(1)$ as $\sigma\to 0$. Then, for small σ

$$\delta(y,\sigma)=\sigma^2\left(\frac{1}{8}\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{5/2}}-\frac{5}{64}\frac{1}{\left(y+\frac{3}{8}+\sigma^2\right)^{7/2}}+\ldots\right.$$
$$\left.\ldots+(-1)^s a_{s-1}\frac{2Q_{s-1}\left(y,\sigma^2\right)}{\left(y+\frac{3}{8}+\sigma^2\right)^{s-3/2}}+\mathcal{O}(1)\right), \quad (36)$$

and therefore $\delta\left(y,\sigma\right)=\mathcal{O}\left(\sigma^2\right)$ as $\sigma\to 0$.

In conclusion, noting that $\mathcal{I}_0$ is a smooth function with derivative $\mathcal{I}_0'$, (33) and (34) yield

$$\mathcal{I}_\sigma(E\{f_\sigma(z)|y,\sigma\})=\mathcal{I}_0(E\{f_\sigma(z)|y,\sigma\})-\sigma^2+$$
$$+\mathcal{O}(\mathcal{I}_0'(E\{f_\sigma(z)|y,\sigma\}))\,\mathcal{O}(\delta(y,\sigma)),$$

where the orders for $\delta\left(y,\sigma\right)$ found above apply.

Equation (29) shows also that $E\{f_\sigma(z)|y,\sigma\}$ is large only if y or σ are large, i.e., $E\{f_\sigma(z)|y,\sigma\}\to+\infty$ if and only if $\breve{y}=y+\frac{3}{8}+\sigma^2\to+\infty$. Therefore, the above analysis implies that $\mathcal{I}_\sigma(D)$ approaches $\mathcal{I}_0(D)-\sigma^2$ for large D as well as for large or small σ. We make this statement more precise in the next section.

4) Direct asymptotics with respect to D and σ: Let $D=E\{f_\sigma(z)|y,\sigma\}$. For large $\breve{y}$, we see from (28) that $\breve{y}\sim D^2$. Hence, (35) yields $\delta(y,\sigma)=\mathcal{O}\left(\sigma^2 D^{-5}\right)$. Since the derivative $\mathcal{I}_0'(D)$ approaches $\frac{D}{2}$ for large D, we then have

$$y=\mathcal{I}_\sigma(D)=\mathcal{I}_0(D)-\sigma^2+\mathcal{O}\left(\sigma^2 D^{-4}\right). \quad (37)$$

For a given fixed σ (and thus considering large y), this becomes

$$y=\mathcal{I}_\sigma(D)=\mathcal{I}_0(D)-\sigma^2+\mathcal{O}\left(D^{-4}\right), \quad (38)$$

whereas for a fixed y (and thus considering large σ), we have $\sigma^2\sim\breve{y}\sim D^2$ and, consequently,

$$\begin{aligned} y=\mathcal{I}_\sigma(D) &= \mathcal{I}_0(D)-\sigma^2+\mathcal{O}\left(D^{-2}\right) \\ &= \mathcal{I}_0(D)-\sigma^2+\mathcal{O}\left(\sigma^{-2}\right). \end{aligned} \quad (39)$$

Finally, for fixed y and $\sigma\to 0$, we have

$$y=\mathcal{I}_\sigma(D)=\mathcal{I}_0(D)-\sigma^2+\mathcal{O}\left(\sigma^2\right). \quad (40)$$

Note that there is no analogous equation for the case $\sigma\neq 0$ and D approaching $E\{f_\sigma(z)|0,\sigma\}$ (i.e., small y), as for this case $\mathcal{I}_\sigma(D)-\mathcal{I}_0(D)-\sigma^2$ converges to a number that, although quite small (as shown by (24) in Section III-F), is typically non-zero. This notwithstanding, (39) and (40) are valid also for $y=0$.

B. Derivation of the ML inverse

To prove (14), let us consider two cases. First, if $D\geq E\left\{f_\sigma\left(z\right)\mid 0,\sigma\right\}$, (13) can be maximized by choosing y in such a way that the maximum of the probability density function (PDF) $p\left(D\mid y\right)$ coincides with D. In other words, $y=\mathcal{I}_\sigma\left(D\right)$ is such that $E\left\{f_\sigma(z)|0,\sigma\right\}=D$, and therefore it maximizes $p\left(D|y\right)$.

Second, if $D\leq E\left\{f_\sigma(z)|0,\sigma\right\}$, then $D-E\left\{f_\sigma(z)|y,\sigma\right\}\leq 0$ for all $y\in[0,+\infty)$. Because the mapping $y\mapsto D-E\left\{f_\sigma(z)|y,\sigma\right\}$ is strictly decreasing, and because variations of y correspond to translations of the PDF, $p\left(D|y\right)=u_0\left(D-E\left\{f_\sigma(z)|y,\sigma\right\}\right)$ is maximized by taking the smallest value of y, i.e., $y=0$, thus giving $\mathcal{I}_{\mathrm{ML}}(D)=0$.

Note that this proof is essentially the same as in [7] for the pure Poisson case, and is presented here for completeness.