MULTI-RESOLUTION AITCHISON GEOMETRY IMAGE DENOISING FOR LOW-LIGHT PHOTOGRAPHY

Thesis

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Electrical Engineering

By

Sarah Victoria Miller

Dayton, Ohio

August, 2020

University *of* Dayton

MULTI-RESOLUTION AITCHISON GEOMETRY IMAGE DENOISING FOR LOW-LIGHT PHOTOGRAPHY

Name: Miller, Sarah Victoria

APPROVED BY:

Keigo Hirakawa, Ph.D.
Advisory Committee Chairman
Associate Professor, Electrical and Computer Engineering

Brad Ratliff., Ph.D.
Committee Member
Associate Professor, Electrical and Computer Engineering

Vijayan K. Asari, Ph.D.
Committee Member
Professor, Electrical and Computer Engineering

Robert J. Wilkens, Ph.D., P.E.
Associate Dean for Research and Innovation
Professor
School of Engineering

Eddy M. Rojas, Ph.D., M.A., P.E.
Dean, School of Engineering

ABSTRACT

MULTI-RESOLUTION AITCHISON GEOMETRY IMAGE DENOISING FOR LOW-LIGHT PHOTOGRAPHY

Name: Miller, Sarah Victoria
University of Dayton

Advisor: Dr. Keigo Hirakawa

In low-photon imaging regime, noise in image sensors are dominated by shot noise, best modeled statistically as Poisson. In this work, we show that the Poisson likelihood function is very well matched with the Bayesian estimation of the "difference of log of contrast of pixel intensities." More specifically, our work takes root in statistical compositional data analysis, whereby we reinterpret the Aitchison geometry as a multiresolution analysis in log-pixel domain. We demonstrate that the difference-log-contrast has wavelet-like properties that correspond well with human visual system, while being robust to illumination variations. We derive a denoising technique based on an approximate conjugate prior for the latent Aitchison variable that gives rise to an explicit minimum mean squared error estimation. The resulting denoising techniques preserves image contrast details that are arguably more meaningful to human vision than the pixel intensity values themselves.

For Dad

ACKNOWLEDGMENTS

I would like to thank NHK for funding this work, Keigo and my lab mates for all of their support, my family, and Chris. I could not have done this without you guys.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Poisson noise—commonly referred to as "photon noise" or "shot noise"—is the predominant source of noise in images not taken in low-light conditions [1]. The heteroskedastic nature of the photon arrival process can be modeled by the following probability:

$$P(X = k) = \frac{e^{\lambda t}(\lambda t)^k}{k!}, \tag{1.1}$$

where X is the number of photons measured in time t and λ is the intensity of the photons [1]. There are several existing algorithms aimed at reducing Poisson noise prevalent in pixel sensors [2] [3] [4]. While these algorithms perform well for low to moderate levels of Poisson noise, the performance of the denoising deteriorates as the signal level decreases.

We propose a novel image denoising method that reinterprets the Aitchison geometry [5] as a multi-resolution analysis in log-pixel domain. Aitchison geometry models "difference of log of pixel contrast" with wavelet-like properties that correspond well with the human visual system. The minimum mean squared error (MMSE) estimation of Aitchison geometry image is comparable or superior to the current state of the art denoising method in terms of mean squared error and structural similarity index in our synthetic image experiments. Visual assessment of the denoised real images shows that the performance of Aitchison-based denoising was most consistent across widely diverse image contents such as textures, weak edges, highlights and dark regions, and smooth gradients.

1.1 Prior Art

Fueled by the ubiquity of imaging devices, image denoising has been an active research topic. Small form factors such as smartphones require aggressive denoising strategy because they lack ability to collect sufficient light. Even large format cameras, such as ones used in broadcasting, are subject to substantial amount of noise under low light conditions in high framerate, high resolution videography. We briefly review representative methods below. Block-matching and 3D filtering (BM3D) method applies a shrinkage operator to a "shape-adaptive" three dimensional discrete cosine transform that has been shown to preserve image features such as edges and textures well [4]. BM3D is by now a *de facto* benchmarking method for image denoising, and its Gaussian-Poisson variant is particularly useful for real-world low-light imaging applications [6].

The distribution of Haar wavelet coefficients of inhomogeneous Poisson process (such as real-world images) is said to be Skellam [7]. Hence Skellam-based denoising methods are designed to estimate the mean Skellam value (i.e. clean wavelet coefficient) as a proxy for estimating the latent pixel intensity—subsequent inverse wavelet transform recovers denoised pixels [2, 8–11]. Skellam minimum risk shrinkage operator (Skellam-MRSO) in particular provably achieves minimum attainable L2 error, but without explicit prior density function [2].

Multiscale multiplicative innovation (MMI) transform is an alternative to Haar wavelet transform, where the local spatial singularities are modeled as local pixel intensity contrast instead of pixel intensity differences [3, 12–14]. Poisson count variables when conditioned on MMI representation of the latent pixel contrast may be reinterpreted as binomial distribution. Methods to estimate MMI coefficients have yielded superior preservation of image details, particularly in low signal levels. MRSO has been extended to MMI, achieving provably minimum attainable L2 error in image denoising [3].

In recent years, deep learning approaches to image denoising have attracted attention. Denoising convolutional neural network (DnCNN) is a feed-forward convolutional neural network leveraging discriminiative model learning [15]. Its hidden layers are designed to remove the latent clean image, yielding residual noise that subsequently may be used to estimate the signal. Fast and flexible denoising convolutional neural network (FFDNet) handles heteroskedastic noise with a single network framework. Learning-based methods have shown promise in "blind" image denoising (denoising without predetermined noise variance) [16, 17].

CHAPTER II

IMAGE SIGNAL AS MULTI-RESOLUTION AITCHISON GEOMETRY

2.1 Background: Aitchison Geometry

In statistics, "compositional data" refers to the notion of modeling proportions as a result of random partitioning, or relative sizes of individual groups compared to the size of all groups combined. For instance, suppose $x \in \mathbb{R}^+$ and $y \in \mathbb{R}^+$ represent the random partitioning of a population $\ell := x + y$. Then the compositional representation of this partitioning is

$$\begin{aligned}(\bar{x}, \bar{y}) :=& \left(\frac{x}{x+y}, \frac{y}{x+y}\right)\\ (x, y) =& \ell \cdot (\bar{x}, \bar{y}),\end{aligned} \tag{2.1}$$

where $\bar{x}$ and $\bar{y}$ is a simplex:

$$1 = \bar{x} + \bar{y}. \tag{2.2}$$

Aitchison pioneered a linear algebraic representation of compositional data $(\bar{x}, \bar{y})$ using logarithms [5]. For instance, define $a, b \in \mathbb{R}$ to be linear combinations of $\log \bar{x}$ and $\log \bar{y}$, as follows:

$$\begin{cases} a := & \log \bar{x} - \log \bar{y} \\ b := & \log \bar{x} + \log \bar{y}. \end{cases} \tag{2.3}$$

Then it is possible to recover $(\bar{x}, \bar{y})$ from a single Aitchison variable a only:

$$(\bar{x}, \bar{y}) = \left(e^{\frac{b+a}{2}}, e^{\frac{b-a}{2}}\right) = e^{\frac{b}{2}}(e^{\frac{a}{2}}, e^{\frac{-a}{2}}) = \frac{(e^{\frac{a}{2}}, e^{-\frac{a}{2}})}{e^{\frac{a}{2}} + e^{-\frac{a}{2}}}. \tag{2.4}$$

Consequently, (x, y) can be reconstructed from (ℓ, a):

$$(x, y) = \ell \cdot \frac{(e^{\frac{a}{2}}, e^{-\frac{a}{2}})}{e^{\frac{a}{2}} + e^{-\frac{a}{2}}}. \tag{2.5}$$

The connection of Aitchison variable a to Poisson count variables will be made explicit in Section A.

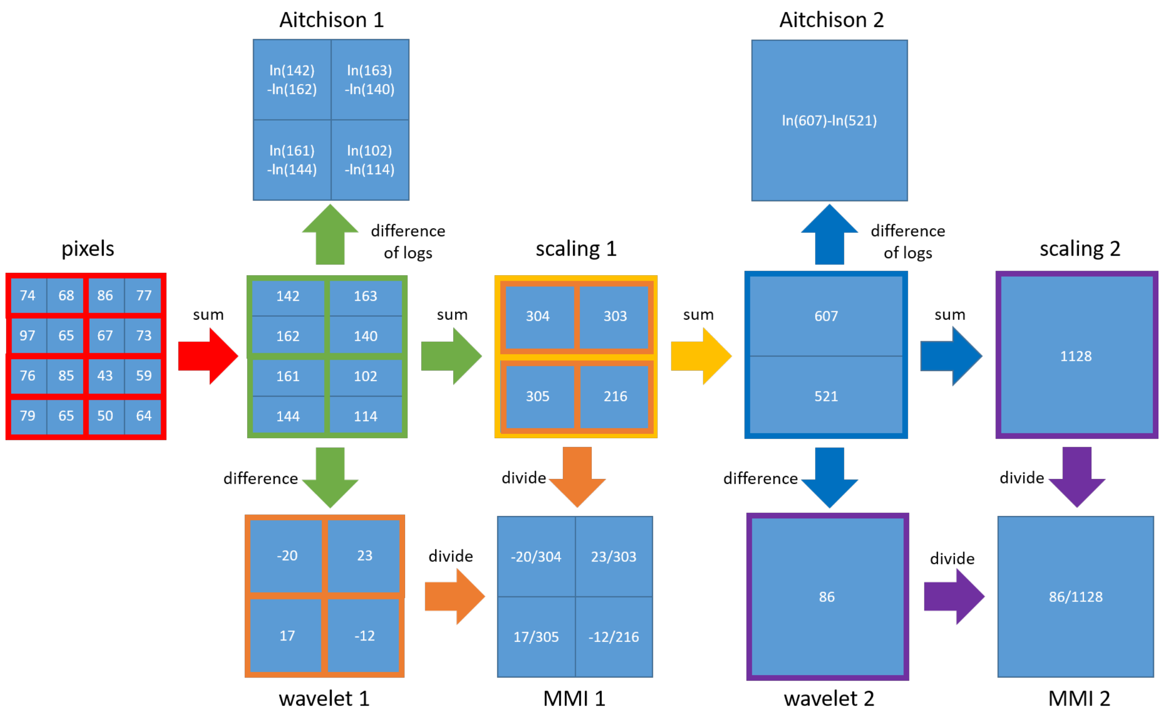


Figure 2.1: A toy example, illustrating 2 level Haar wavelet, MMI, and Aitchison decomposition using a 4×4 image. Colors of the arrow indicate the grouping of pixels that the operations are applied to. The horizontal orientation decomposition shown in this diagram generalizes to vertical and diagonal decompositions in a straightforward manner.

2.2 Multi-Resolution Image Signal Representation

Suppose $x, y \in \mathbb{R}^+$ represent neighboring pixel intensities. Then we may also regard Aitchison variable as the difference of log pixel intensities:

$$\begin{aligned} a &= \log\left(\frac{x}{x+y}\right) - \log\left(\frac{y}{x+y}\right) \\ &= \log(x) - \log(y). \end{aligned} \tag{2.6}$$

Alternatively, under the classical image formation models, (x, y) represent a camera capturing light reflecting on a surface:

$$(x, y) = i \cdot (r_x, r_y). \tag{2.7}$$

Here, $r_x, r_y \in [0,1]$ represent reflectance values of x, y, respectively; and $i \in \mathbb{R}^+$ is the illumination irradiance that is constant or slowly-varying over space. Under this classical model, Aitchison variable $a \in \mathbb{R}$ in (2.3) is a quantity intrinsic to the scene reflectance and invariant to illumination:

$$\begin{aligned} a &= \log\left(\frac{i \cdot r_x}{i \cdot r_x + i \cdot r_y}\right) - \log\left(\frac{i \cdot r_y}{i \cdot r_x + i \cdot r_y}\right) \\ &= \log(r_x) - \log(r_y). \end{aligned} \tag{2.8}$$

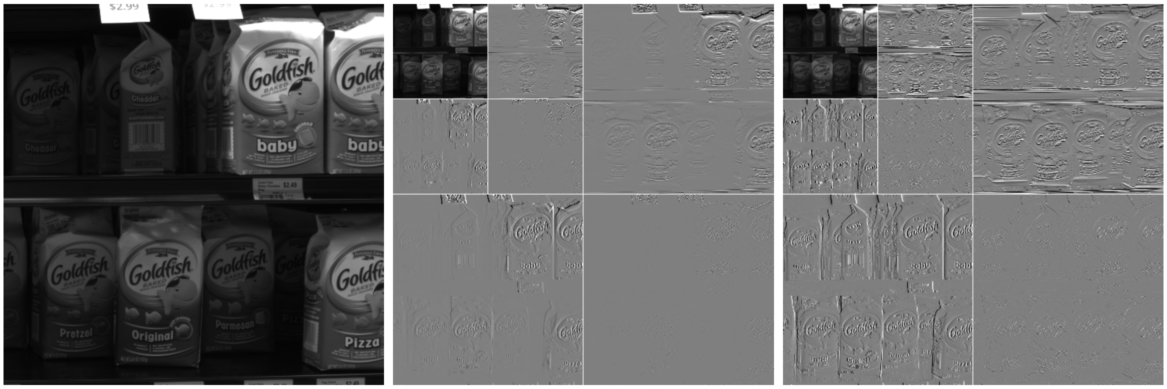
Hence, we conclude that Aitchison variable $a \in \mathbb{R}$ is a stable representation of the scene reflectance across all illumination levels, particularly well-matched for low-light or high dynamic range imaging applications.

Recall (2.5). In this paper, we propose to estimate ℓ and a from the observed noisy pixel values as a "proxy denoising task" for estimating the latent pixel intensity values x and y. We propose an Aitchison geometry representation of a 2D image that is inspired in part by Haar wavelet transform [18] as well as multiplicative multiscale innovation (MMI) transform [3, 12–14]. Multi-resolution Haar wavelet, MMI, and Aitcheson multiscale transforms are illustrated in Figure 2.1. Note that all three decompositions can be reversed—meaning the pixels can be recovered from the wavelet/MMI/Aitchison coefficients (along with the scaling coefficients). The classical Haar wavelet and scaling coefficients are computed as sums and differences of pixels. In the context of Poisson noise problems, the distribution of noisy Haar wavelet coefficients are modeled as Skellam [19]; and many Skellam denoising (Skellam mean estimation) techniques have been developed [2, 19–23]. However, owing to the complicated coupling of latent variables to Poisson and Skellam distributions, Haar wavelet-based image denoising techniques are largely limited to univariate estimation.

MMI coefficients are ratios computed by normalizing the wavelet coefficients by the scaling coefficients. Intuitively, MMI coefficients quantify contrast between neighboring pixels (differences between pixels relative to the pixel magnitudes) and they are invariant to scene illumination levels. Estimation of MMI coefficients from noisy pixel observations (which have binomial likelihood function) have been shown to preserve edge structures better than Skellam denoising methods [3, 12–14]. However, similarly to the Poisson and Skellam distributions, ratios are very complicated to work with, especially in the context of multivariate MMI denoising. Thus existing MMI-based denoising methods are largely univariate methods.

Multi-resolution Aitchison variables are computed as differences of log pixel magnitudes. Like Haar wavelet transform, the horizontal, vertical, and diagonal orientations can be computed at different resolutions (Figure 2.1 shows horizontal coefficients). Like MMI, the compositional data is invariant to scene illumination levels, and describe intensity patterns in relative terms. But unlike Haar wavelet-based Skellam mean estimation or MMI based binomial denoising, Aitchison denoising is highly amenable to multivariate estimation, thanks to the fact that multivariate normal distributions are approximate conjugate priors for the Aitchison likelihood functions (proven in Theorem 2 below). The proposed Aitchison denoising greatly improves the preservation of image features, even though the multivariate denoising design is relatively straightforward.

Figure 2.2 shows examples of Haar wavelet and Aitchison coefficients for a relatively clean image. These transform coefficients share several attributes—the transform coefficients are sparse, and the dominant edges are very clearly represented. Differences are more noticeable when comparing bright and dark regions of the image, however. In Haar wavelets representation, the edges and textures in bright regions of the image are disproportionately more pronounced than similar details in dark regions of the image. This inconsistency arises due to the nonuniform illumination over the scene. Image denoising without regard for local



(a) (b) (c)

Figure 2.2: Example multi-resolution Haar wavelet and Aitchison decomposition. (a) Original image. (b) Haar wavelet decomposition. (c) Aitchison coefficients. The spatial details captured by Haar wavelet coefficients are influenced by the illumination. By contrast, the edge and texture details represented by the Aitchison coefficients are consistent throughout specular and shadow regions.

illumination levels often leads to oversmoothing in dark regions and undersmoothing the well-lit regions. By comparison, the compositional Aitchison representation in Figure 2.2 exhibits image details that are stable under spatially varying illumination conditions. This underscores the potential of Aitchison-based denoising algorithms to preserving texture and edge structures in bright and dark regions of the image, equally.

Finally, Aitchison decomposition illustrated in Figure 2.1 is compatible with the notion of overcomplete transform, made popular in wavelet-based denoising methods. Specifically, the Aitchison decomposition in Figure 2.1 is not shift-invariant due to the downsampling occurring as a result of combining two pixels at a time. The Aitchison representation can be made shift-invariant if we repeat the Aitchison decomposition in Figure 2.1 by spatially shifting the input image. Overcomplete transforms are known to increase denoising performance and reduce denoising artifacts. See [24, 25] for further explanation.

CHAPTER III

IMAGE DENOISING

3.1 Estimation in Aitchison Geometry

Let $x \in \mathbb{R}$ and $y \in \mathbb{R}$ denote two pixel intensity values. The corresponding observed pixel values $X \in \mathbb{Z}$ and $Y \in \mathbb{Z}$ are independent Poisson random variables:

$$X|x \sim Poiss(x), \qquad Y|y \sim Poiss(y). \tag{3.1}$$

Suppose we rewrite them as sum $S \in \mathbb{Z}$ and difference $W \in \mathbb{Z}$, as follows:

$$\begin{cases} S = X + Y \\ W = X - Y, \end{cases} \qquad \begin{cases} X = \frac{S+W}{2} \\ Y = \frac{S-W}{2}. \end{cases} \tag{3.2}$$

Then, we may equivalently rewrite the likelihood function in (3.1) as

$$P[X, Y|x, y] = P[S, W|\ell, a] = P[W|\ell, a, S]P[S|\ell, a]. \tag{3.3}$$

The theorem below makes the Aitchison likelihood function in (3.3) explicit.

Theorem 1 (Aitchison Likelihood)**.**

$$P[S, W|\ell, a] = P[W|a, S]P[S|\ell] \tag{3.4}$$

where $S|\ell \sim Poiss(\ell)$ *and*

$$P[W|a, S] = \frac{S!}{\left(\frac{S+W}{2}\right)!\left(\frac{S-W}{2}\right)!} \exp\left(\frac{Wa}{2} - S\log\left(e^{\frac{a}{2}} + e^{-\frac{a}{2}}\right)\right). \tag{3.5}$$

Proof is provided in Appendix A. Theorem 1 can be generalized to multivariate Aitchison likelihood. Specifically, denote by $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{Z}^N$ and $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{W}, \boldsymbol{\ell}, \boldsymbol{a} \in \mathbb{R}^N$ groups of pixels and corresponding coefficients. Furthermore, assuming the statistical independence of noise[1]:

$$P[\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{x},\boldsymbol{y}] = P[\boldsymbol{X}|\boldsymbol{x}]P[\boldsymbol{Y}|\boldsymbol{y}] = \prod_{n=1}^{N} P[X_n|x_n]P[Y_n|y_n], \tag{3.6}$$

the multivariate likelihood function in (3.6) may be reparameterized as the multivariate conditional independence of Aitchison variable:

$$P[\boldsymbol{S},\boldsymbol{W}|\boldsymbol{\ell},\boldsymbol{a}] = P[\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}]P[\boldsymbol{S}|\boldsymbol{\ell}] = \prod_{n=1}^{N} P[W_n|a_n,S_n]P[S_n|\ell_n]. \tag{3.7}$$

The significance of Theorem 1 is that, if ℓ and a were statistically independent (recall Section 2.2), the minimum mean squared error (MMSE) estimation of a and ℓ can be decoupled completely, simplifying the denoising process. Below, we develop a building block for multivariate Aitchison denoising.

Theorem 2 (Aitchison Conjugate Prior)**.** *Suppose the Aitchison variable $\boldsymbol{a} \in \mathbb{R}^N$ is distributed as multivariate normal random variable:*

$$\boldsymbol{a} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}), \tag{3.8}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ is a covariance matrix, and $\mathbf{0} \in \mathbb{R}^N$ is a zero vector. Suppose further that $\boldsymbol{\ell}$ is independent of $\boldsymbol{a}$. Then the posterior density of $\boldsymbol{a}$ is approximately normal:

$$\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S} \overset{approx}{\sim} \mathcal{N}\left(\frac{(\boldsymbol{\Sigma}^{-1}+\frac{1}{4}\operatorname{diag}(\boldsymbol{S}))^{-1}\boldsymbol{W}}{2}, (\boldsymbol{\Sigma}^{-1}+\frac{1}{4}\operatorname{diag}(\boldsymbol{S}))^{-1}\right) \tag{3.9}$$

Corollary 1 (Aitchison MMSE)**.** *Suppose the hypotheses of Theorem 2 hold. Then the minimum mean squared error (MMSE) estimate of the latent variable $\boldsymbol{a}$ is the posterior mean of the form:*

$$\mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S}] \approx \frac{(\boldsymbol{\Sigma}^{-1}+\frac{1}{4}\operatorname{diag}(\boldsymbol{S}))^{-1}\boldsymbol{W}}{2}. \tag{3.10}$$

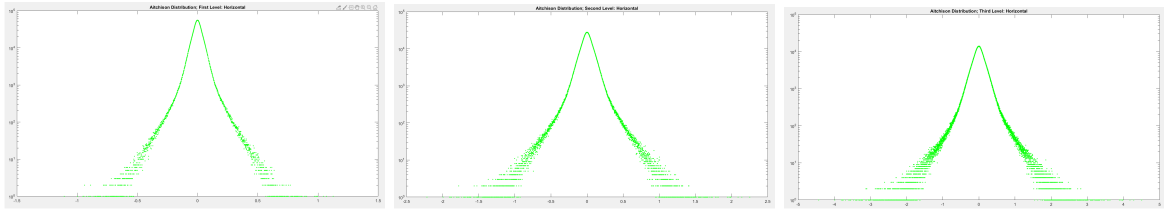
with expected mean squared error performance of $(\boldsymbol{\Sigma}^{-1}+\frac{1}{4}\operatorname{diag}(\boldsymbol{S}))^{-1}$.

[1] Conditional independence in (3.6) and (3.7) should *not* be misinterpreted as statistical independence of signal $\boldsymbol{x}$ and $\boldsymbol{y}$, observations $\boldsymbol{X}$ and $\boldsymbol{Y}$, nor Aitchison variables $\boldsymbol{a}$.

Proof is found in Appendix B. Corollary 1 follows directly from Theorem 2 and the classical Bayesian estimation theory. **We emphasize that Theorem 2 should *not* be confused with the notions of "normal approximation" commonly used to simplify Poisson or binomial distributions**. The latter refers to the approximation to the likelihood functions for the noisy observations $\boldsymbol{X}$ or $\boldsymbol{W}$—it is a special case of the central limit theorem, whereby the approximation becomes poorer as the signal decreases (i.e. less applicable to low-light imaging). Instead, what *is* being claimed by Theorem 2 is that the normal distribution is an approximate conjugate prior of Aitchison likelihood. The approximation holds regardless of the signal strength $\boldsymbol{S}$, favorable to the low-light imaging conditions we are concerned with (when the central limit theorem does not apply). We emphasize the unusual nature of approximating the posterior of the latent variable $\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S}$—unlike the more conventional normal approximations designed to simplify the likelihood function $\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}$. In this regard, Theorem 2 is perhaps more similar in spirit to variational Bayes approaches [26–28].

3.2 Image Denoising

Recalling Section 2.2 and Figure 2.1, multi-resolution Aitchison geometry models local variations of the neighboring pixels or scaling coefficients $\boldsymbol{x}$ and $\boldsymbol{y}$ at different scales and orientations. Estimation of Aitchison variables $\boldsymbol{a}$ at every scale and orientation from noisy scaling coefficients $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a "proxy denoising task" for estimating $\boldsymbol{x}$ and $\boldsymbol{y}$. Figure 3.1 shows empirical distribution of Aitchison variables at levels 1 through 3. It is evident from these plots that the distribution of Aitchison variable is heavier tail than the normal density in Theorem 2. Drawing inspirations from [29], we represent the heavy tail distribution as a Gaussian mixture.



(a) (b) (c)

Figure 3.1: Estimation of Aitchison variable $\boldsymbol{a}$ at levels 1 through 3 for horizontal orientation. (a) 1st level. (b) 2nd level. (c) 3rd level.

Theorem 3 (Gaussian Mixture Modeling)**.** *Suppose the distribution of Aitchison variable* $\boldsymbol{a} \in \mathbb{R}^N$ *is described by the Gaussian mixture model:*

$$\boldsymbol{a}|z \sim \mathcal{N}(\boldsymbol{0}, z\boldsymbol{\Sigma}) \tag{3.11}$$

where $z \sim P[z]$ *is a hyperparameter independent of* $\boldsymbol{S}$*. Let*

$$\boldsymbol{M}_z^{-1} = (z\boldsymbol{\Sigma})^{-1} + \frac{1}{4}\operatorname{diag}(\boldsymbol{S}). \tag{3.12}$$

Then the minimum mean squared error (MMSE) estimate of the latent variable $\boldsymbol{a}$ *is:*

$$\mathbb{E}[\boldsymbol{a}|\boldsymbol{W}, \boldsymbol{S}] \approx \frac{\int \frac{\boldsymbol{M}_z\boldsymbol{W}}{(2\pi)^{\frac{N}{2}}|\det(4\boldsymbol{M}_z^{-1})|^{\frac{1}{2}}} \exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8}\right) P[z]dz}{2\int \frac{1}{(2\pi)^{\frac{N}{2}}|\det(4\boldsymbol{M}_z^{-1})|^{\frac{1}{2}}} \exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8}\right) P[z]dz}. \tag{3.13}$$

The proof of Theorem 3 is found in Appendix C. Even though (3.13) resembles standard normal density equations, we caution the readers that it should not be confused with the additive white Gaussian noise case examined in [29] or its variant using the normal approximation of Poisson likelihood function—$\boldsymbol{M}_z$ matrix in (3.12) playing the role of the covariance matrix is specific to estimation in the Aitchison geometry.

For imaging, (3.13) is carried out at every pixel location, including the matrix inverses. The matrix $\boldsymbol{M}_z$ cannot be precomputed because it is data-dependent (involving $\boldsymbol{S}$). Hence the complexity of literal execution of the multivariate estimation technique in Theorem 3 is high. Has shown by Theorem 4 below, however, Theorem 3 can be accelerated with a minor assumption.

Theorem 4 (Fast Aitchison Denoising)**.** *Let*

$$\boldsymbol{\Sigma} = \boldsymbol{Q}^T \boldsymbol{\Lambda} \boldsymbol{Q}. \tag{3.14}$$

denote an eigen-decomposition of $\boldsymbol{\Sigma}$ *with* $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$*, and*

$$\begin{aligned} \boldsymbol{V} = (v_1, \ldots, v_N)^T :=& \boldsymbol{Q}\boldsymbol{W} \in \mathbb{R}^N \\ \boldsymbol{u} = (u_1, \ldots, u_N)^T :=& \boldsymbol{Q}\boldsymbol{a} \in \mathbb{R}^N. \end{aligned} \tag{3.15}$$

Suppose that the following approximation holds:

$$\mathrm{diag}(\boldsymbol{S}) = \bar{S}\boldsymbol{I}, \tag{3.16}$$

where $\bar{S} = \sum_{n=1}^{N} S_n/N$ *is an average value. Then the MMSE of* $\boldsymbol{a}$ *in Theorem 3 may be rewritten as*

$$\mathbb{E}[u_n|\boldsymbol{W},\boldsymbol{S}] \approx \frac{\int \frac{2z\lambda_n}{\bar{S}z\lambda_n+4} V_n \frac{\exp\left(-\frac{1}{8}\sum_{n=1}^{N} V_n^2 \frac{4z\lambda_n}{\bar{S}z\lambda_n+4}\right)}{\sqrt{\prod_{n=1}^{N}(2\pi)(4z^{-1}\lambda_n+\bar{S})}} P[z]dz}{\int \frac{\exp\left(-\frac{1}{8}\sum_{n=1}^{N} V_n^2 \frac{4z\lambda_n}{\bar{S}z\lambda_n+4}\right)}{\sqrt{\prod_{n=1}^{N}(2\pi)(4z^{-1}\lambda_n+\bar{S})}} P[z]dz}, \tag{3.17}$$

$$\mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S}] = \boldsymbol{Q}^T \mathbb{E}[\boldsymbol{u}|\boldsymbol{W},\boldsymbol{S}]. \tag{3.18}$$

The proof of the theorem is found in Appendix D. Note that the estimation technique in Theorem 4 is significantly less complex than Theorem 3. Specifically, eigen decomposition in (3.14) can be precomputed so $\boldsymbol{Q}$ is known *a priori*. Moreover, the eigen-transformed value $\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{W}$ need to be computed only once, independent of the z value. The inversion of data-dependent matrix is replaced by scalar divisions, which are not only more efficient than matrix inversion, but can be precomputed as a look-up-table.

Drawing inspirations from [29], we estimate the denoising parameter $\boldsymbol{\Sigma}$ from noisy data. Consider the following theorem.

Theorem 5 (Covariance Estimation). *Let $\boldsymbol{\Sigma} = \mathrm{Cov}\{\boldsymbol{a}\} := \mathbb{E}[\boldsymbol{a}\boldsymbol{a}^T] \in \mathbb{R}^{N\times N}$ denote covariance matrix of Aitchison variable $\boldsymbol{a}$. Then*

$$\mathrm{Cov}\{\boldsymbol{W}\} \approx \mathbb{E}[S]\left(\mathbb{E}\left[\frac{4}{(e^{-\frac{a}{2}}+e^{-\frac{a}{2}})^2}\right]\boldsymbol{I} + \frac{1}{4}\boldsymbol{\Sigma}\right), \tag{3.19}$$

where $S \in \mathbb{R}$ and $a \in \mathbb{R}$ denote all scaling and Aitchison variables in the image, respectively.

Proof is found in Appendix E. In practice, $\mathbb{E}\left[\frac{4}{(e^{-\frac{a}{2}}+e^{-\frac{a}{2}})^2}\right] \approx 4$ is a stable number across all image types. Hence, we arrive at our estimate of covariance matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} \approx \frac{4}{\mathbb{E}[S]}\,\mathrm{Cov}\{\boldsymbol{W}\} - 4\boldsymbol{I} \tag{3.20}$$

In our implementation, we set $P[z]$ to be discrete random variables. That is, we assume $z \in \{z_1, \ldots, z_K\}$ where

$$P[z = z_k] = \pi_k \tag{3.21}$$

and $\sum_k \pi_k = 1$. In this case, the estimation in Theorem 4 becomes

$$\mathbb{E}[u_n|\boldsymbol{W},\boldsymbol{S}] \approx \frac{\sum_k \frac{2z_k\lambda_n}{\bar{S}z_k\lambda_n+4} V_n \frac{\exp\left(-\frac{1}{8}\sum_{n=1}^N V_n^2 \frac{4z_k\lambda_n}{\bar{S}z_k\lambda_n+4}\right)}{\sqrt{\prod_{n=1}^N (2\pi)(4z_k^{-1}\lambda_n+\bar{S})}}\pi_k}{\sum_k \frac{\exp\left(-\frac{1}{8}\sum_{n=1}^N V_n^2 \frac{4z_k\lambda_n}{\bar{S}z_k\lambda_n+4}\right)}{\sqrt{\prod_{n=1}^N (2\pi)(4z_k^{-1}\lambda_n+\bar{S})}}\pi_k}. \tag{3.22}$$

We used empirically determined choices of hyperparameters $\{z_1, \ldots, z_K\}$ and their corresponding probabilities $\{\pi_1, \ldots, \pi_K\}$ tabulated in Table 3.1. They worked well across many types of images in our experiments.

Table 3.1: Hyperparameters, z_k, and corresponding probabilities, π_k

	z_1	z_2	z_3	π_1	π_2	π_3
4th level	46.9807	4.69807	0.00469807	0.33	0.33	0.33
3rd level	11.7595	1.17595	0.00117595	0.33	0.33	0.33
2nd level	.29417	.029417	2.9417e-05	0.33	0.33	0.33
1st level	0.073566	0.0073566	7.3566e-06	0.33	0.33	0.33

CHAPTER IV

RESULTS

Images used in this paper are taken by Sony α7Riii, a 42 megapixel camera. All images are taken in raw sensor format, using the "pixel shift" function that allows demosaicking to be bypassed (so as to obtain a full pixel sample at each pixel location with the highest spatial fidelity). Whenever possible, all camera parameters were set manually so as to maintain consistency across shots. We took a series of clean images in well-lit environments with long exposures. We also took a series of noisy images in very low-light conditions with short exposure times and/or with neutral density filters. In addition, noisy observation of X-Rite ColorChecker was captured to measure the noise statistics.

To construct a greyscale image from noisy raw sensor data, we apply affine transform to red, green, and blue channels such that the noise statistics match that of Poisson—noise variance equals signal intensity. This process is often referred to as "Poisson correction" in the literature, and it is similar to the process of finding the noise variance in additive white Gaussian noise problems. Note that additive or signal-independent noise (such as dark current) can be modeled by Poisson correction also (see [30] for explanation; mathematically identical to Poisson-Gaussian noise treatment in [6]). We then sum together the affine-transformed red, green, and blue channels, yielding a single greyscale Poisson image (sum of Poisson is Poisson). Example images are shown in Figure 4.1.

The clean images are used for synthetic experiments as a proxy for noise-free images. The pseudo-random noise we add to the clean image mimic noisy raw sensor statistics—the signal range, the black offset, noise variance as a function of signal intensity, dark current noise variance match the characteristics we measured in noisy raw sensor images. Specifically, we apply affine transform to red, green, and blue channels, matching the black

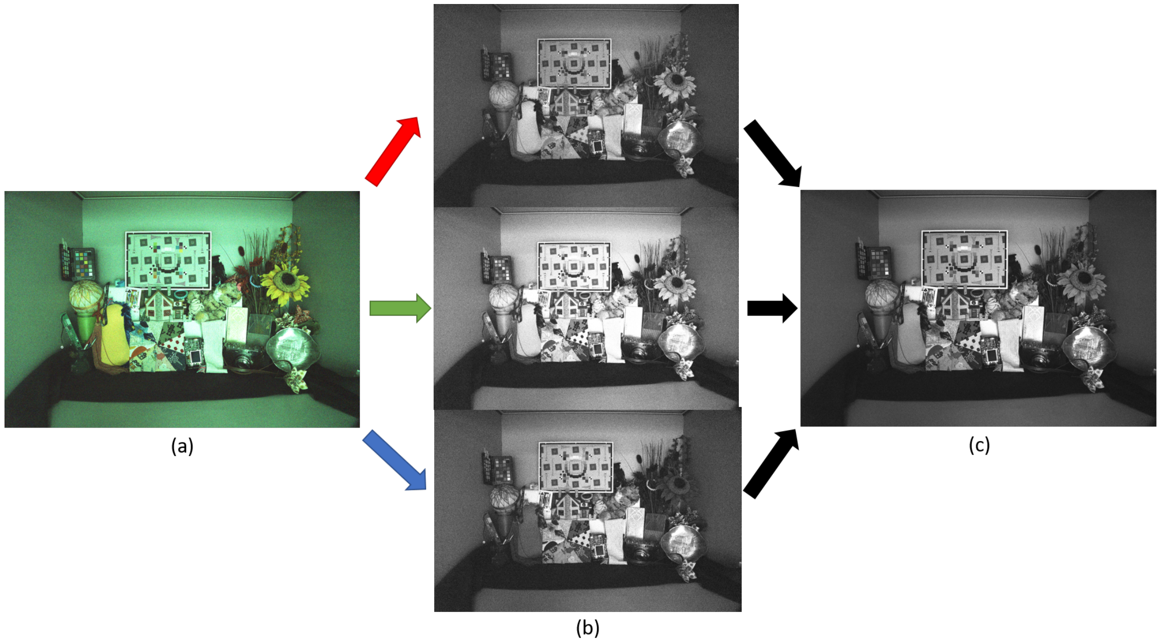


Figure 4.1: Construction of a greyscale image. (a) Input RGB image. (b) Separate, affine transformed red, green, and blue color channels. (c) Resulting greyscale image.

offset and the dynamic range of noisy Poisson corrected images. Summing the color channels yields noise-free reference greyscale image. We generate a pseudo-random Poisson image from this greyscale image, to which we apply denoising algorithms.

In our implementation, we used 4 levels of overcomplete multi-resolution Aitchison geometry decomposition (recall horizontal, vertical, and diagonal subbands per level). The performance of the proposed image denoising method is compared to state-of-the-art and standard benchmark denoising methods. They are BM3D [4] with Poisson-Gaussian correction [6], Skellam MRSO [2], MMI MRSO [3], DnCNN [15], and FFDNet [31].

4.1 Synthetic Experiment

The purpose of the synthetic experiment is to assess the denoising performance quantitatively. Table 4.1 shows the peak signal-to-noise-ratio (PSNR) and structural similarity index metric (SSIM) computed in the pixel domain. According to PSNR, the MMI MRSO and Skellam MRSO outperform all methods by more than 1dB. However, ranking by SSIM confirms the superiority of the proposed Aitchison denoising method in perceptual error metrics.

Recall the proposed method is a minimum mean square error estimator of Aitchison variable that is a lighting-invariant representation. To this end, we also benchmark in Table 4.1 the denosing methods by the reconstruction quality of Aitchison variable. We provide MSE instead of PSNR, because there is no nominal "peak" value in Aitchison domain. In addition, we adopt multiscale SSIM to compute the reference and recovered Aitchison variables at each subband [32]—we report the MS-SSIM score, averaged over the various Aitchison subbands. It is clear from Table 4.1 that the proposed method outperforms other denoising methods in Aitchison variable reconstruction.

4.2 Real Noisy Sensor Data Experiment

In absence of reference or clean image, we assess denoising experiment with real noisy sensor qualitatively. As shown by Figures 4.2-4.5, the proposed Aitchison denoising method effectively reduces Poisson noise while preserving fine details that improves overall visual experience. In particular, the denoising of smooth regions of the image have flat appearance with little residual noise (e.g. the body of the camera in Figure 4.3 and the card in Figure 4.4). We compare this to MMI MRSO, Skellam MRSO, and FFDNet, where the under-

Table 4.1: Denoising performance of various methods, averaged over 25 images. In addition to PSNR and SSIM computed in pixel domain, we compute MSE and multiscale SSIM of Aitchison variables (MSE instead of PSNR because there is no nominal "peak" value in Aitchison domain).

	Pixel Domain		Aitchison Domain	
	PSNR	SSIM	MSE	MS-SSIM
Noisy	20.2839	0.3032	1.6486e-02	0.3960
BM3D [4]	33.7488	0.9381	2.7668e-04	0.7626
Skellam MRSO [2]	**34.8017**	0.9297	2.9350e-04	0.7623
MMI MRSO [3]	34.7872	0.9188	3.3848e-04	0.7505
DnCNN [15]	31.9346	0.7760	1.0894e-03	0.6020
FFDNet [31]	31.1129	0.9223	3.8014e-04	0.7380
Aitchison (Proposed)	34.3041	**0.9402**	**2.5898e-04**	**0.7751**

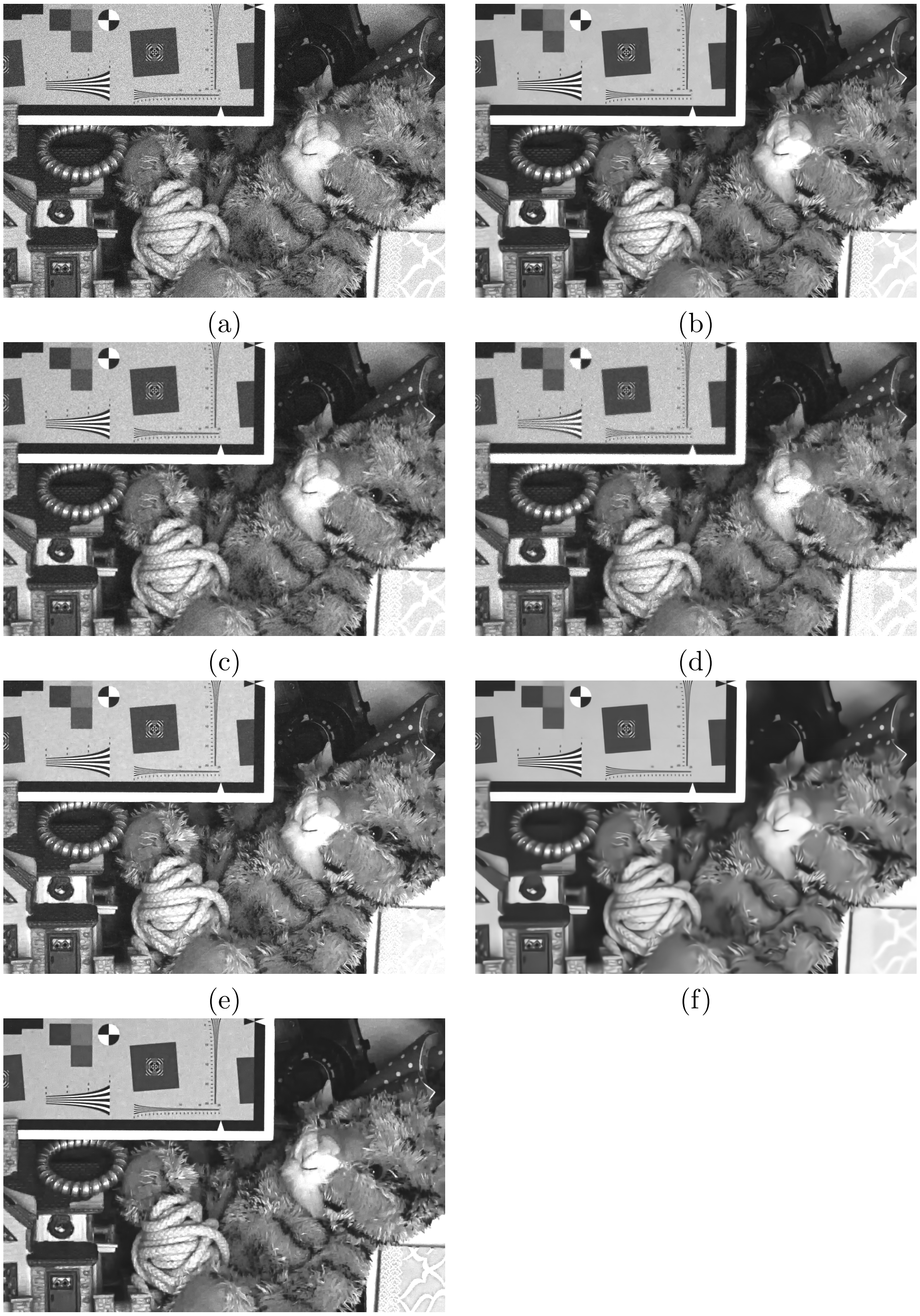
aggressive denoising fails to suppress residual noise. BM3D and DnCNN on the other hand introduce false textures formed by reinforcing the structures formed by the instantiation of noise (in smooth gradients such as flat region next to the playing card in Figure 4.4).

In a lightly textured regions (e.g. the matte-surface resolution chart in Figure 4.2, or fabric in Figure 4.5), the Aitchison denoising method has a tendency to under- or over-smooth (depending on the texture content), though visually it is superior to DnCNN and FFDNet. Thanks to adaptive shape transform, BM3D is superior at preserving repetitive texture pattern, though false texturing is evidenced in other textured regions (e.g. matte resolution chart in Figure 4.2, fabrics next to camera in Figure 4.3).

The textures behind the camera in Figure 4.3 is a patterning on a tin can, exhibiting a very high dynamic range (juxtaposition of highlights and medium pixel intensities). Here, illumination invariance of Aitchison helps preserve the extreme contrast, while the brightest pixels blew out and dominated over the darker pixels in all other methods.

Strong edges and textures are preserved well by all methods, including the Aitchison, though qualitatively they yield different characteristics and appearances—and as a result, the appearance has been "altered" compared to the noisy input image. For instance, the hair of stuffed animal in Figure 4.2 is slightly blurred in Skellam-MRSO, and slightly weaker in contrast in MMI-MRSO, and smeared in FFDNet. Though geometrically accurate, BM3D artificially seems to enhances the texture. DnCNN and Aitchison seems most similar to the noisy image in visual impact of the textures of the tiger fur (without blurring or enhancing contrast).

Weak edges (e.g. house siding in Figure 4.4) are difficult to preserve. BM3D yielded a textured artifacts while Skellam-MRSO and MMI-MRSO underdenoised and FFDNet oversmoothed. DnCNN and the Aitchison denoising are most similar, with former having more artifacts and the latter being slightly noisier.



(a) (b) (c) (d) (e) (f) (g)

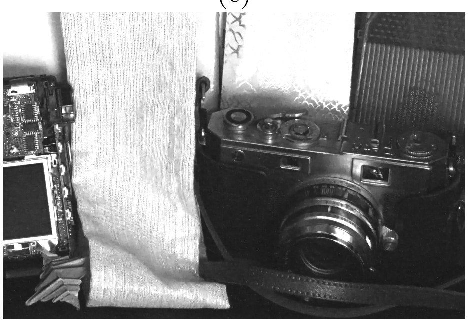
Figure 4.2: (a) Original image. (b) BM3D denoised result. (c) Skellam denoised result. (d) MMI denoised result. (e) DnCNN denoised result. (f) FFDNet denoised result. (g) Aitchison denoised result (proposed).

(a)

(b)

(c)

(d)

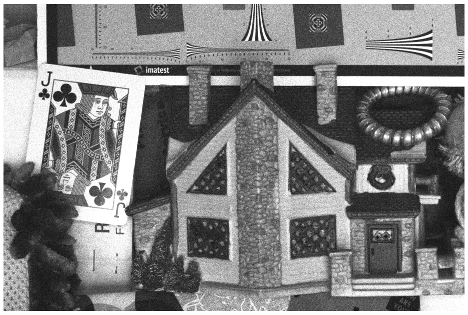


(e)

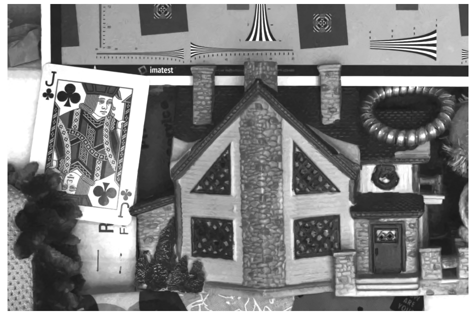
(f)

(g)

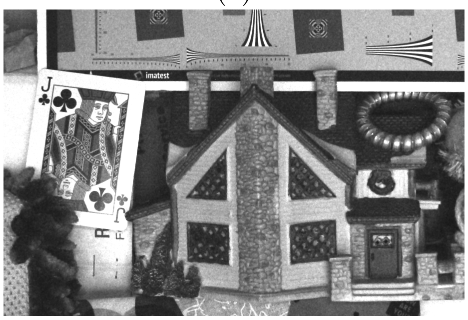
Figure 4.3: (a) Original image. (b) BM3D denoised result. (c) Skellam denoised result. (d) MMI denoised result. (e) DnCNN denoised result. (f) FFDNet denoised result. (g) Aitchison denoised result (proposed).



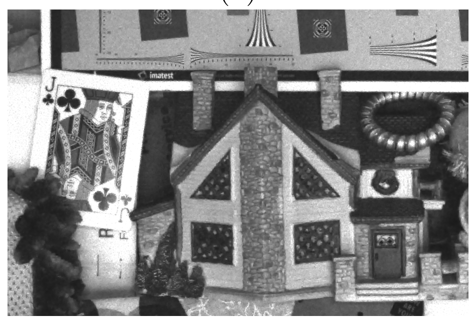
(a)



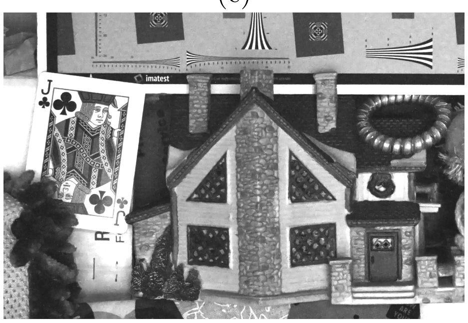
(b)



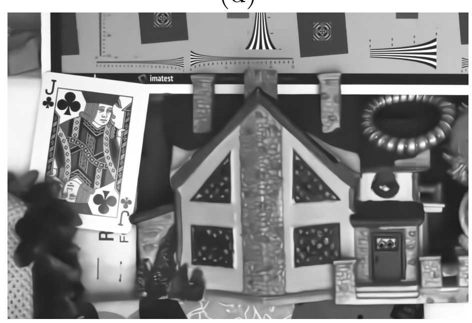
(c)



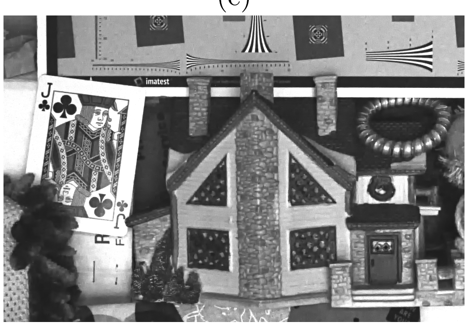
(d)



(e)

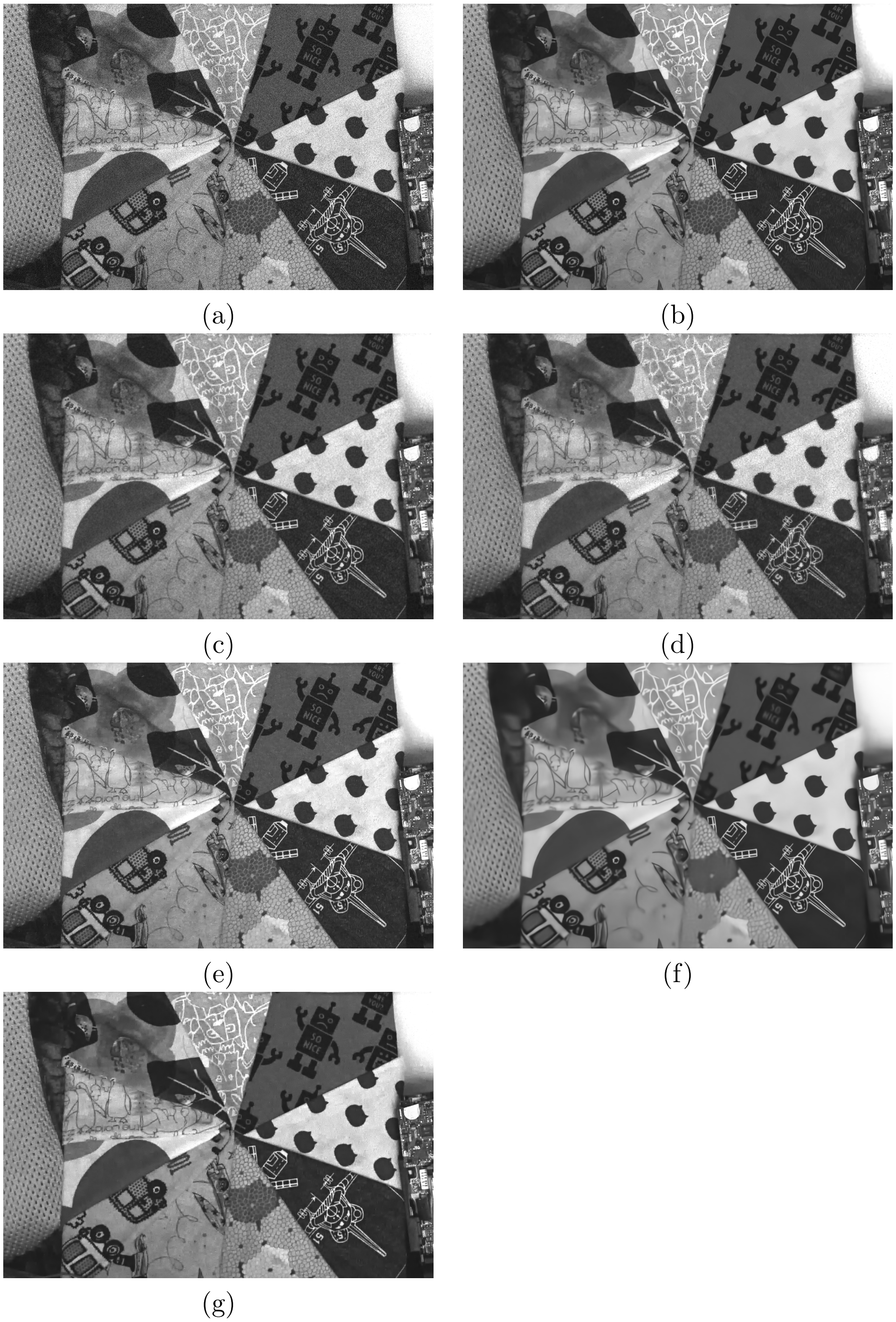


(f)



(g)

Figure 4.4: (a) Original image. (b) BM3D denoised result. (c) Skellam denoised result. (d) MMI denoised result. (e) DnCNN denoised result. (f) FFDNet denoised result. (g) Aitchison denoised result (proposed).



(a) (b) (c) (d) (e) (f) (g)

Figure 4.5: (a) Original image. (b) BM3D denoised result. (c) Skellam denoised result. (d) MMI denoised result. (e) DnCNN denoised result. (f) FFDNet denoised result. (g) Aitchison denoised result (proposed).

CHAPTER V

CONCLUSION

In this paper, we presented a novel image denoising method based on Aitchison geometry. We proposed a Haar wavelet-like Aitchison multi-resolution representation of image signal reflectance stable across all illumination levels. We explicitly derived the Aitchison likelihood function, and found approximate conjugate priors. We developed a multivariate minimum mean squared error estimation of Aitchison variable using mixture of approximate conjugate priors to match the heavy-tailed nature of the latent Aitchison variable. Simulation experiments confirmed that the proposed Aitchison-based denoising is competitive or better than the state-of-the-art methods in pixel recovery. When denoising performance is measured by the accuracy of Aitchison variable reconstruction, our method is superior. Visual assessment of real noisy sensor data shows that—while no method (state-of-the-art as well as ours) is perfect—the denoising performance was most consistent among the methods compared, across widely diverse image contents such as textures, weak edges, highlights and dark regions, and smooth gradients.

BIBLIOGRAPHY

[1] S. W. Hasinoff, “Photon, poisson noise,” in *Computer Vision.* Springer US, 2014, pp. 608–610.

[2] W. Cheng and K. Hirakawa, “Minimum risk wavelet shrinkage operator for poisson image denoising,” *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1660–1671, may 2015.

[3] ——, “Towards optimal denoising of image contrast,” *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3446–3458, jul 2018.

[4] 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.

[5] J. Aitchison, *The Statistical Analysis of Compositional Data.* Dordrecht: Springer Netherlands, 1986.

[6] M. Makitalo and A. Foi, “Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise,” *IEEE transactions on image processing*, vol. 22, no. 1, pp. 91–103, 2012.

[7] J. G. Skellam, “The frequency distribution of the difference between two poisson variates belonging to different populations,” *Journal of the Royal Statistical Society: Series A*, vol. 109, p. 296, 1946.

[8] K. Hirakawa and P. J. Wolfe, “Efficient multivariate skellam shrinkage for denoising photon-limited image data: An empirical bayes approach,” in *2009 16th IEEE International Conference on Image Processing (ICIP).* IEEE, 2009, pp. 2961–2964.

[9] ——, “Skellam shrinkage: Wavelet-based intensity estimation for inhomogeneous poisson data,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1080–1093, 2011.

[10] F. Luisier, T. Blu, and M. Unser, “Image denoising in mixed poisson–gaussian noise,” *IEEE Transactions on image processing*, vol. 20, no. 3, pp. 696–708, 2010.

[11] Y. Hwang, J.-S. Kim, and I. S. Kweon, “Difference-based image noise modeling using skellam distribution,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 7, pp. 1329–1341, 2011.

[12] K. E. Timmermann and R. D. Nowak, “Multiscale modeling and estimation of poisson processes with application to photon-limited imaging,” *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, 1999.

[13] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for poisson inverse problems," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1811–1825, 2000.

[14] E. D. Kolaczyk, "Bayesian multiscale models for poisson processes," *Journal of the American Statistical Association*, vol. 94, no. 447, pp. 920–933, 1999.

[15] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, jul 2017.

[16] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview."

[17] C. Tian, Y. Xu, L. Fei, and K. Yan, "Deep learning for image denoising: A survey."

[18] K. H. Talukder and K. Harada, "Haar wavelet based approach for image compression and quality assessment of compressed image."

[19] M. Alrefaya and H. Sahli, "A novel adaptive probabilistic nonlinear denoising approach for enhancing PET data sinogram," *Journal of Applied Mathematics*, vol. 2013, pp. 1–14, 2013.

[20] L. Zhao, J. Zhang, and Z. Wei, "Skellam distribution based adaptive two-stage non-local methods for photon-limited poisson noisy image reconstruction," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, sep 2017.

[21] K. Hirakawa and P. J. Wolfe, "Efficient multivariate skellam shrinkage for denoising photon-limited image data: An empirical bayes approach," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, nov 2009.

[22] ——, "Skellam shrinkage: Wavelet-based intensity estimation for inhomogeneous poisson data," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1080–1093, feb 2012.

[23] Y. Hwang, J.-S. Kim, and I. S. Kweon, "Difference-based image noise modeling using skellam distribution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1329–1341, jul 2012.

[24] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *IEEE Transactions on image processing*, vol. 17, no. 8, pp. 1342–1352, 2008.

[25] Xin Li and M. T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, vol. 3, 2000, pp. 300–303 vol.3.

[26] M. Titsias and M. Lázaro-Gredilla, "Doubly stochastic variational bayes for non-conjugate inference," in *International conference on machine learning*, 2014, pp. 1971–1979.

[27] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-iid noise structure," *IEEE transactions on cybernetics*, vol. 48, no. 3, pp. 1054–1066, 2017.

[28] B. Keng, "Variational bayes and the mean-field approximation," 2017.

[29] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, nov 2003.

[30] S. Yang and B.-U. Lee, "Poisson-gaussian noise reduction using the hidden markov model in contourlet domain for fluorescence microscopy images," *PLOS ONE*, vol. 10, no. 9, p. e0136964, sep 2015.

[31] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, sep 2018.

[32] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[33] H. S. Wall, *Analytic theory of continued fractions.* Bronx, N.Y: Chelsea, 1973.

APPENDIX A

Proof of Theorem 1

It is a well known fact that the sum of Poisson variables are Poisson

$$S|x, y \sim Poiss(x + y). \tag{A.1}$$

Hence, $P[S|x, y] = P[S|\ell, a] = P[S|\ell]$. In addition, Poisson variables are conditionally binomial, in the following sense:

$$p[X, Y|x, y, S] = \frac{S!}{X!Y!} \bar{x}^X \bar{y}^Y. \tag{A.2}$$

We use Aitchison geometry in (2.3) to decompose (A.2) further:

$$\begin{aligned} p[X, Y|x, y, S] =& \frac{S!}{X!Y!} \exp\left(X \log \bar{x} + Y \log \bar{y}\right) \\ =& \frac{S!}{\left(\frac{S+W}{2}\right)! \left(\frac{S-W}{2}\right)!} \exp\left(\left(\frac{S+W}{2}\right) \log \bar{x} + \left(\frac{S-W}{2}\right) \log \bar{y}\right) \\ =& \frac{S!}{\left(\frac{S+W}{2}\right)! \left(\frac{S-W}{2}\right)!} \exp\left(\frac{Sb + Wa}{2}\right). \end{aligned} \tag{A.3}$$

Recalling (2.4), we may rewrite $s \in \mathbb{R}$ as

$$\begin{aligned} b =& \log \bar{x} + \log \bar{y} = \log\left(\frac{e^{a/2}}{e^{a/2} + e^{-a/2}}\right) + \log\left(\frac{e^{-a/2}}{e^{a/2} + e^{-a/2}}\right) \\ =& \frac{a}{2} - \log\left(e^{a/2} + e^{-a/2}\right) + \frac{-a}{2} - \log\left(e^{a/2} + e^{-a/2}\right) \\ =& -2 \log\left(e^{a/2} + e^{-a/2}\right). \end{aligned} \tag{A.4}$$

Combining, we arrive at the Aitchison likelihood function:

$$\begin{aligned} P[X, Y|x, y, S] =& \frac{S!}{\left(\frac{S+W}{2}\right)! \left(\frac{S-W}{2}\right)!} \exp\left(\frac{Wa}{2} - S \log\left(e^{a/2} + e^{-a/2}\right)\right) \\ =& P[W|a, S]. \end{aligned} \tag{A.5}$$

APPENDIX B

Proof of Theorem 2

. The posterior density function $\boldsymbol{a}$ is

$$P[\boldsymbol{a}|\boldsymbol{W}, \boldsymbol{S}] = \frac{P[\boldsymbol{W}|\boldsymbol{a}, \boldsymbol{S}]P[\boldsymbol{a}|\boldsymbol{S}]}{P[\boldsymbol{W}|\boldsymbol{S}]}. \tag{B.1}$$

Since magnitude $\boldsymbol{\ell}$ is independent of Aitchison $\boldsymbol{a}$, $\boldsymbol{S}$ is independent of $\boldsymbol{a}$. That is, $P[\boldsymbol{a}|\boldsymbol{S}] = P[\boldsymbol{a}]$. Expanding the numerator,

$$\begin{aligned} P[\boldsymbol{W}|\boldsymbol{a}, \boldsymbol{S}]P[\boldsymbol{a}|\boldsymbol{S}] =& \frac{\prod_{n=1}^{N} \frac{S_n!}{(\frac{S_n+W_n}{2})!(\frac{S_n-W_n}{2})!}}{(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{1}{2}}} \\ & \times \exp\left(-\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}{2} + \frac{\boldsymbol{W}^T\boldsymbol{a}}{2} - \sum_{n=1}^{N} S_n \log\left(e^{\frac{a_n}{2}} + e^{-\frac{a_n}{2}}\right)\right). \end{aligned} \tag{B.2}$$

Substituting Taylor series expansion of $\log(e^{\frac{a_n}{2}} + e^{-\frac{a_n}{2}})$:

$$\log(e^{\frac{a_n}{2}} + e^{-\frac{a_n}{2}}) \approx \log(2) + \frac{a_n^2}{8}, \tag{B.3}$$

the numerator in (B.2) may be approximated as

$$\begin{aligned} P[\boldsymbol{W}|\boldsymbol{a}, \boldsymbol{S}]P[\boldsymbol{a}|\boldsymbol{S}] \approx& \frac{\prod_{n=1}^{N} \frac{S_n!}{(\frac{S_n+W_n}{2})!(\frac{S_n-W_n}{2})!}}{(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{1}{2}}} \\ & \times \exp\left(-\frac{\boldsymbol{a}^T\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{4}\operatorname{diag}(\boldsymbol{S})\right)\boldsymbol{a} + \boldsymbol{W}^T\boldsymbol{a}}{2} - \boldsymbol{S}^T\log\boldsymbol{2}\right). \end{aligned} \tag{B.4}$$

Letting $\boldsymbol{M}^{-1} = \boldsymbol{\Sigma}^{-1} + \frac{1}{4}\operatorname{diag}(\boldsymbol{S})$ and completing the square yields

$$\begin{aligned} P[\boldsymbol{W}|\boldsymbol{a}, \boldsymbol{S}]P[\boldsymbol{a}|\boldsymbol{S}] \approx& \frac{\prod_{n=1}^{N} \frac{S_n!}{(\frac{S_n+W_n}{2})!(\frac{S_n-W_n}{2})!}}{(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{1}{2}}} \\ & \times \exp\left(-\frac{(\boldsymbol{a} - \frac{\boldsymbol{MW}}{2})^T\boldsymbol{M}^{-1}(\boldsymbol{a} - \frac{\boldsymbol{MW}}{2}) - \frac{\boldsymbol{W}^T\boldsymbol{MW}}{4}}{2} - \boldsymbol{S}^T\log\boldsymbol{2}\right). \end{aligned} \tag{B.5}$$

Or equivalently, posterior density is approximately normal:

$$\boldsymbol{a}|\boldsymbol{W}, \boldsymbol{S} \sim \mathcal{N}\left(\frac{\boldsymbol{MW}}{2}, \boldsymbol{M}\right) \tag{B.6}$$

As a side-note, the empirical distribution of $\boldsymbol{a}$ is highly concentrated around zero, meaning the Taylor series expansion in (B.3) holds very well in practice.

APPENDIX C

Proof of Theorem 3

By a property known as *total expectation*, we have:

$$\mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S}] = \int \mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S},z]P[z|\boldsymbol{W},\boldsymbol{S}]dz. \tag{C.1}$$

Thanks to Theorem 2, we may compute $\mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S},z]$ as

$$\mathbb{E}[\boldsymbol{a}|\boldsymbol{W},\boldsymbol{S},z] \approx \left((z\boldsymbol{\Sigma})^{-1} + \frac{1}{4}\operatorname{diag}(\boldsymbol{S})\right)^{-1}\boldsymbol{W}. \tag{C.2}$$

On the other hand, the posterior hyperparameter has the following distribution:

$$P[z|\boldsymbol{W},\boldsymbol{S}] = \frac{P[\boldsymbol{W}|\boldsymbol{S},z]P[z|\boldsymbol{S}]}{\int P[\boldsymbol{W}|\boldsymbol{S},z]P[z|\boldsymbol{S}]dz} \tag{C.3}$$

Letting $\boldsymbol{M}_z^{-1} = (z\boldsymbol{\Sigma})^{-1} + \frac{1}{4}\operatorname{diag}(\boldsymbol{S})$ and substituting (B.5)

$$\begin{aligned}
P[\boldsymbol{W}|\boldsymbol{S},z] &= \int P[\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S},z]P[\boldsymbol{a}|\boldsymbol{S},z]d\boldsymbol{a} \\
&\approx \int \frac{\prod_{n=1}^{N}\frac{S_n!}{(\frac{S_n+W_n}{2})!(\frac{S_n-W_n}{2})!}}{(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{1}{2}}}\exp\left(-\frac{(\boldsymbol{a}-\frac{\boldsymbol{M}_z\boldsymbol{W}}{2})^T\boldsymbol{M}_z^{-1}(\boldsymbol{a}-\frac{\boldsymbol{M}_z\boldsymbol{W}}{2}) - \frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{4}}{2} - \boldsymbol{S}^T\log\boldsymbol{2}\right)d\boldsymbol{a} \\
&= \frac{\prod_{n=1}^{N}\frac{S_n!}{(\frac{S_n+W_n}{2})!(\frac{S_n-W_n}{2})!}}{(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{\Sigma})|^{\frac{1}{2}}}(2\pi)^{\frac{N}{2}}|\det(\boldsymbol{M}_z)|^{\frac{1}{2}}\exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8} - \boldsymbol{S}^T\log\boldsymbol{2}\right),
\end{aligned} \tag{C.4}$$

we simplify (C.3) to:

$$P[z|\boldsymbol{W},\boldsymbol{S}] = \frac{\frac{1}{(2\pi)^{\frac{N}{2}}|\det(4\boldsymbol{M}_z^{-1})|^{\frac{1}{2}}}\exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8}\right)P[z|\boldsymbol{S}]}{\int\frac{1}{(2\pi)^{\frac{N}{2}}|\det(4\boldsymbol{M}_z^{-1})|^{\frac{1}{2}}}\exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8}\right)P[z|\boldsymbol{S}]dz}. \tag{C.5}$$

Finally, recalling statistical independence of z and $\boldsymbol{S}$, we have

$$P[z|\boldsymbol{S}] = P[z]. \tag{C.6}$$

Substituting (C.2), (C.5), and (C.6) to (C.1) proves the theorem.

APPENDIX D

Proof of Theorem 4

The eigen-vectors in $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ are also eigen-vectors of $\boldsymbol{M}$:

$$\begin{aligned} \boldsymbol{Q}\boldsymbol{M}_z^{-1}\boldsymbol{Q}^T =& \boldsymbol{Q}\left((z\boldsymbol{\Sigma})^{-1} + \frac{\bar{S}\boldsymbol{I}}{4}\right)\boldsymbol{Q}^T \\ =& z^{-1}\boldsymbol{\Lambda}^{-1} + \frac{\bar{S}\boldsymbol{I}}{4}. \end{aligned} \tag{D.1}$$

Then we may simplify the numerator and denominator of (3.13) as by the relation:

$$\begin{aligned} \frac{\exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{M}_z\boldsymbol{W}}{8}\right)}{(2\pi)^{\frac{N}{2}}|\det(4\boldsymbol{M}_z^{-1})|^{\frac{1}{2}}} =& \frac{\exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{Q}^T\left(z^{-1}\boldsymbol{\Lambda}^{-1}+\frac{\bar{S}\boldsymbol{I}}{4}\right)^{-1}\boldsymbol{Q}\boldsymbol{W}}{8}\right)}{(2\pi)^{\frac{N}{2}}|\det(4z^{-1}\boldsymbol{\Lambda}^{-1}+\bar{S}\boldsymbol{I})|^{\frac{1}{2}}} \\ =& \frac{\exp\left(-\frac{1}{8}\boldsymbol{V}^T\begin{bmatrix}\frac{1}{z\lambda_1}+\frac{\bar{S}}{4} & & \\ & \ddots & \\ & & \frac{1}{z\lambda_N}+\frac{\bar{S}}{4}\end{bmatrix}^{-1}\boldsymbol{V}\right)}{(2\pi)^{\frac{N}{2}}\prod_{n=1}^N(4z^{-1}\lambda_n+\bar{S})^{\frac{1}{2}}} \\ =& \frac{\exp\left(-\sum_{n=1}^N\frac{1}{8}V_n^2\frac{4z\lambda_n}{\bar{S}z\lambda_n+4}\right)}{\sqrt{\prod_{n=1}^N 2\pi(4z^{-1}\lambda_n+\bar{S})}}. \end{aligned} \tag{D.2}$$

That is, V_n and $V_{n'}$ are independent when $n \neq n'$. Similarly, we have the following relationship:

$$\boldsymbol{M}_z\boldsymbol{W} = \boldsymbol{Q}^T\begin{bmatrix}\frac{4z\lambda_1}{\bar{S}z\lambda_1+4} & & \\ & \ddots & \\ & & \frac{4z\lambda_N}{\bar{S}z\lambda_N+4}\end{bmatrix}\boldsymbol{V} \tag{D.3}$$

Substituting (D.2) and (D.3) into (3.13) in Theorem 3 proves the theorem.

APPENDIX E

Proof of Theorem 5

Consider the Taylor series expansion of $\tanh(\cdot)$ function

$$\tanh\left(\frac{a}{2}\right) = \frac{a}{2} - \frac{a^3}{24} + \frac{a^5}{240} + \dots \tag{E.1}$$

It is evident that this Taylor expansion converges very quickly. Another series

$$\tanh\left(\frac{a}{2}\right) = \frac{a}{2 + \frac{a^2}{6 + \frac{a^2}{10 + \frac{a^2}{14 + \dots}}}} \tag{E.2}$$

converges even faster [33]. In Figure E.1, we compare $\tanh(a/2)$ and their approximations, for their real-world $a \in \mathbb{R}$ values computed from a typical clean image. One can see that the errors of second order approximations ($\tanh(a/2) \approx a/2 - a^3/24$ or $\tanh(a/2) \approx \frac{a}{2+a^2/6}$) are practically negligible. Moreover, the linear approximation ($\tanh(a/2) \approx a/2$) is adequate for majority of Aitchison variables—the error contributes to sharpening the image, in fact.

Recall total covariance theorem:

$$\mathrm{Cov}\{\boldsymbol{W}\} = \mathbb{E}[\mathrm{Cov}\{\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}\}] + \mathrm{Cov}\{\mathbb{E}[\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}]\}. \tag{E.3}$$

Owing to the independence of noise, the covariance matrix of $\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}$ is diagonal

$$\begin{aligned}\mathrm{Cov}\{\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}\} =& 4\,\mathrm{diag}\left(S_1 \frac{e^{\frac{a_1}{2}}}{e^{\frac{a_1}{2}} + e^{-\frac{a_1}{2}}} \frac{e^{-\frac{a_1}{2}}}{e^{-\frac{a_1}{2}} + e^{-\frac{a_1}{2}}}, \dots, S_N \frac{e^{\frac{a_N}{2}}}{e^{\frac{a_N}{2}} + e^{-\frac{a_N}{2}}} \frac{e^{-\frac{a_N}{2}}}{e^{-\frac{a_N}{2}} + e^{-\frac{a_N}{2}}}\right)\\ =& 4\,\mathrm{diag}\left(\frac{S_1}{e^{-\frac{a_1}{2}} + e^{-\frac{a_1}{2}}}, \dots, \frac{S_N}{e^{-\frac{a_N}{2}} + e^{-\frac{a_N}{2}}}\right).\end{aligned} \tag{E.4}$$

Taking its expected value, we solve for the first term in (E.3):

$$\mathbb{E}[\mathrm{Cov}\{\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}\}] = \mathbb{E}[S]\mathbb{E}\left[\frac{4}{(e^{-\frac{a}{2}} + e^{-\frac{a}{2}})^2}\right]\boldsymbol{I}. \tag{E.5}$$

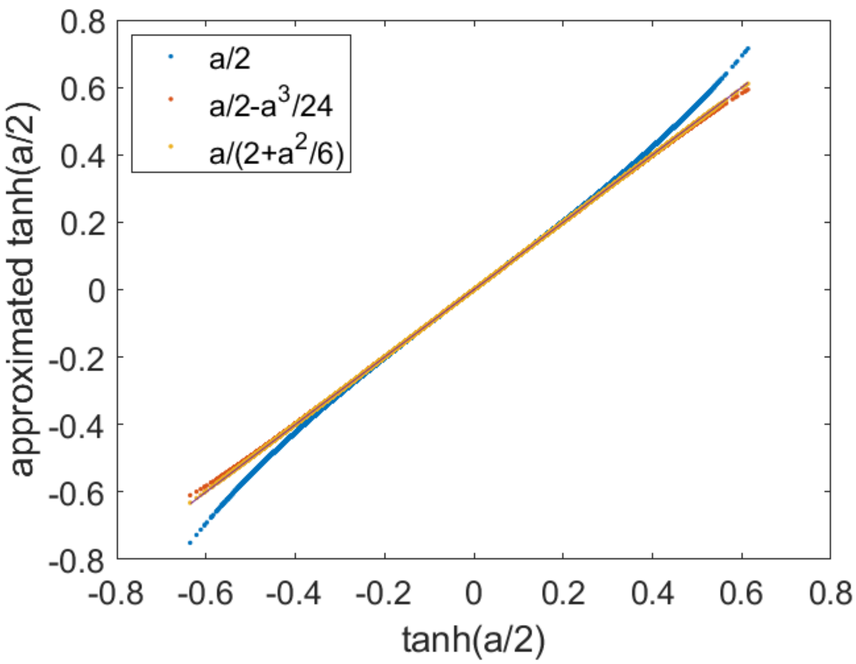


Figure E.1: Approximation of $\tan(a/2)$.

Consider the expected value

$$\mathbb{E}[\boldsymbol{W}|\boldsymbol{a},\boldsymbol{S}] = \mathrm{diag}(\boldsymbol{S})\tanh(\boldsymbol{a}). \tag{E.6}$$

Its covariance matrix is

$$\mathrm{Cov}\{\mathbb{E}[\boldsymbol{W}|\,\mathrm{diag}(\boldsymbol{S})\tanh(\boldsymbol{a})]\} = \mathrm{Cov}\{\mathrm{diag}(\boldsymbol{S})\tanh(\boldsymbol{a})\} \approx \frac{\mathbb{E}[S]\boldsymbol{\Sigma}}{4}, \tag{E.7}$$

where we have used the independence of $\boldsymbol{S}$ and $\boldsymbol{a}$ as well as the Taylor series approximation in B.3 in the last step.