FOURIER AND FILTERBANK ANALYSES OF SIGNAL-DEPENDENT NOISE

Keigo Hirakawa

Harvard University, Department of Statistics
One Oxford Street, Cambridge, MA 02139 USA
hirakawa@stat.harvard.edu

ABSTRACT

Owing to the lack of resolution of the measurement and the randomness inherent in the signal and the measuring devices, the measurement noise is often signal-dependent. Although the statistical modeling of filterbank, wavelets, and short-time Fourier coefficients enjoys immense popularity, transform-based estimation of signal is difficult because the effects of signal-dependent noise permeate across multiple coefficients and subbands. In this work, we show how a general class of signal-dependent noise can be characterized to an arbitrary precision in a Haar filterbank and Fourier representation. The structure of noise in the transform domain admits a variant of Stein's unbiased estimate of risk conducive to processing the corrupted signal in the transform domain, and estimators involving Poisson processes are discussed.

***Index Terms*—** Fourier transform, filterbank, signal-dependent noise, Bayesian estimation, Stein's unbiased estimate of risk.

1. INTRODUCTION

Real-world sensing devices are subject various types of measurement noise. For example, it is well known that the lack of resolution (e.g. quantization), randomness inherent in the signal (e.g. photon/packet arrival), and variabilities in the measuring devices (e.g. thermal noise, electron leakage) contribute to a significant degradation of signal. Estimation of the sequential data $\boldsymbol{f} \in \mathbb{R}^N$ given noisy observations $\boldsymbol{g} \in \mathbb{R}^N$ therefore plays a prominent role in communication, signal processing, imaging, and MRI applications.

To illustrate the challenges in signal estimation problems, suppose we adopt a Bayesian statistics point of view; that is, we model the signals in terms of the prior probability distribution of the latent variable ($p(\boldsymbol{f})$) and the likelihood of the observation conditioned on the latent variable ($p(\boldsymbol{g}|\boldsymbol{f})$). Bayesian statistical estimation and inference techniques make use of the posterior probability, or the probability of the latent variable conditioned on the observation ($p(\boldsymbol{f}|\boldsymbol{g})$), which is proportional to the product of the prior probability distribution of the latent variable and the likelihood function. Motivated by the prior knowledge and empirical studies, statistical modeling of the latent variable in the linear transform domain has enjoyed tremendous popularity—in particular, filterbank, wavelets, and short-time Fourier transforms provide convenient platforms for specifying the prior because their coefficients exhibit temporal and spectral locality, sparsity, and energy compaction properties [1–5].

In this paradigm, the special case of additive white Gaussian noise (AWGN) is studied almost exclusively because the posterior of the transform coefficients is readily accessible when the likelihood function has a closed form in the transform domain. The assumption that the noise is AWGN, however, is inadequate for many real-world applications because the measurement noise is almost always dependent on the rangespace of the signal $\boldsymbol{f}$, effects of which permeate across multiple transform coefficients and subbands. For instance, the number of electrons or photons encountered in a measuring device during an integration period is a Poisson process $g_m|\boldsymbol{f} \overset{\text{i.i.d.}}{\sim} \mathcal{P}(f_m)$ where f_m is the expected electron/photon count per integration period and it is proportional to the electric current or the light intensity [6]. As the integration period increases, $p(g_m|\boldsymbol{f})$ converges weakly to $\mathcal{N}(f_m, f_m)$. In this paper, we consider a more general likelihood model whose conditional variance of $g_m|\boldsymbol{f}$ is a function of its conditional mean, $E[g_m|\boldsymbol{f}]$. That is, $g_m|\boldsymbol{f} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(g_m, h(f_m)^2)$, or

$$g_m = f_m + h(f_m)\,\varepsilon_m, \tag{1}$$

where $\varepsilon_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ is independent of $\boldsymbol{f}$, and $h : \mathbb{R} \to \mathbb{R}$ is the standard deviation of noise as a function of f_m.

One existing strategy to address signal-dependent noise in (1) is to design an invertible nonlinear operator $\gamma : \mathbb{R}^N \to \mathbb{R}^N$ on the observation that (approximately) decouples the signal and noise: $\gamma(\boldsymbol{g})|\gamma(\boldsymbol{f}) \sim \mathcal{N}(\gamma(\boldsymbol{f}), \boldsymbol{I})$ [7–9]. An AWGN-based signal estimation technique is used to estimate $\gamma(\boldsymbol{f})$ given $\gamma(\boldsymbol{g})$, and the inverse transform $\gamma^{-1}(\cdot)$ yields an estimate of $\boldsymbol{f}$. Although this approach modularizes the designs of $\gamma(\cdot)$ and the estimator, the signal model assumed for $\boldsymbol{f}$ no longer holds true for $\gamma(\boldsymbol{f})$ and the optimality of the estimator in the new domain does not translate to optimality in the rangespace of $\boldsymbol{f}$. Alternatively, sufficient smoothness in $\boldsymbol{f}$ and $h(\cdot)$ implies slowly changing noise variance. This motivates a *local* AWGN model, where $h(f_m)^2$ is approximated with a constant over a moving window, and an estimation method designed for AWGN is used to estimate $\boldsymbol{f}$ within this window [10]. This approach is not robust to the singularities in $\boldsymbol{f}$ and noise variance estimation employs heuristics to decouple the noise and the latent variable.

In this paper, we derive a novel and efficient representation of signal-dependent noise in the Haar filterbank (HFT) and Fourier domains with precision up to the Kth moment when $h(\cdot)$ is K-differentiable. Though there exist other filterbank/wavelet transforms with better frequency separation, the advantage to encoding the likelihood function in the transform domain (with asymptotic accuracy) is that the posterior distribution of the latent variables is readily accessible. In light of this, we propose two strategies for manipulating the corrupted signal in the transform domain. First, a maximum *a posteriori* (MAP) estimator of the noise free coefficients is developed from the posterior distribution analysis. Second, the structure of noise in the transform domain admits a variant of Stein's unbiased estimate of risk for a transform-based parametric estimator [11–13].

The rest of this paper is organized as follows. We show useful properties of HFT and Fourier in Section 2. An asymptotic representation of signal-dependent noise is derived in Section 3, and its statistical interpretation are discussed in Section 4 before the concluding remarks in Section 5. Due to page constraints, simulation results are regrettably reserved for future publications.

2. PERMUTATION IN TRANSFORM OPERATORS

In the following discussion, results obtained from the Propositions 2.1 and 2.2 below are subsequently used to prove our main result in the Corollary 2.3. Let $N = 2^n$ be the length of $\boldsymbol{f}$, $\boldsymbol{e}_j \in \mathbb{R}^N$ be the jth standard basis, and $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{f}$, $\boldsymbol{\Phi} \in \mathbb{R}^{N\times N}$ is a linear orthogonal transform. Let $\boldsymbol{\phi}_j \in \mathbb{R}^N$ be the jth column of $\boldsymbol{\Phi}$, and define the point-wise multiplication operator $\odot$ as $\boldsymbol{\phi}_j \odot \boldsymbol{\phi}_{j'} = \mathrm{diag}(\boldsymbol{\phi}_j)\boldsymbol{\phi}_{j'} = \mathrm{diag}(\boldsymbol{\phi}_{j'})\boldsymbol{\phi}_j$.

In the case that $\boldsymbol{\Phi}$ is a HFT matrix, we have

$$\boldsymbol{\Phi} = \underbrace{\boldsymbol{\Phi}_0 \otimes \cdots \otimes \boldsymbol{\Phi}_0}_{n \text{ times}} \qquad \boldsymbol{\Phi}_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

where $\otimes$ is a Kronecker product. Note that $\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}/N$ and $\boldsymbol{\Phi}^T = \boldsymbol{\Phi}$, and ± 1 comprises all entries of $\boldsymbol{\Phi}$. Let $\boldsymbol{\phi}_j = \boldsymbol{\phi}_j^* \in \{\pm 1\}^N$ be the jth column of $\boldsymbol{\Phi}$ and $G = \{\boldsymbol{\phi}_j | \forall j\}$ be a set of all columns of $\boldsymbol{\Phi}$. Then we show that $(G, \odot)$ is an abelian group with some desirable properties.

Proposition 2.1. *(abelian group) Suppose $G = \{\boldsymbol{\phi}_j | \forall j\}$ is a set where $\boldsymbol{\phi}_j$ is the jth column of HFT matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N\times N}$. Then $(G, \odot)$ is an abelian group that is isomorphic to the quotient group $\{\mathbb{Z}/2\mathbb{Z}\}^n$ under addition.*

Proof. Hence forth, let the column index of HFT be represented using binary number $\boldsymbol{j} = (j_1, \cdots, j_n)^T \in \{\mathbb{Z}/2\mathbb{Z}\}^n$, where $\boldsymbol{j}+\boldsymbol{j}' = (j_1+j_1', \cdots, j_n+j_n')^T$. From the definition of $\boldsymbol{\Phi}$,

$$\boldsymbol{\phi}_{\boldsymbol{j}} = \begin{bmatrix} 1 \\ (-1)^{j_n} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ (-1)^{j_2} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ (-1)^{j_1} \end{bmatrix}.$$

Then the following fact emerges:

$$\begin{aligned}
&\boldsymbol{\phi}_{\boldsymbol{j}} \odot \boldsymbol{\phi}_{\boldsymbol{j}'} = \mathrm{diag}(\boldsymbol{\phi}_{\boldsymbol{j}})\boldsymbol{\phi}_{\boldsymbol{j}'} \\
&= \left\{ \begin{bmatrix} 1 & 0 \\ 0 & (\text{-}1)^{j_n} \end{bmatrix} \otimes \mathrm{diag}\left(\begin{bmatrix} 1 \\ (\text{-}1)^{j_{n\text{-}1}} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ (\text{-}1)^{j_1} \end{bmatrix} \right) \right\} \boldsymbol{\phi}_{\boldsymbol{j}'} \\
&= \begin{bmatrix} 1 \\ (\text{-}1)^{j_n + j_n'} \end{bmatrix} \otimes \left\{ \mathrm{diag}\left(\begin{bmatrix} 1 \\ (\text{-}1)^{j_{n\text{-}1}} \end{bmatrix} \otimes \cdots \right) \left(\begin{bmatrix} 1 \\ (\text{-}1)^{j_{n\text{-}1}'} \end{bmatrix} \otimes \cdots \right) \right\}.
\end{aligned}$$

By recursion,

$$\boldsymbol{\phi}_{\boldsymbol{j}} \odot \boldsymbol{\phi}_{\boldsymbol{j}'} = \begin{bmatrix} 1 \\ (-1)^{j_n + j_n'} \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ (-1)^{j_1 + j_1'} \end{bmatrix} = \boldsymbol{\phi}_{\boldsymbol{j}+\boldsymbol{j}'}.$$

This result immediately satisfies the axioms necessary to establish that $(G, \odot)$ is an abelian group (closure, associativity, identity element, inverse element). The bijective map $\boldsymbol{\pi} : G \to \{\mathbb{Z}/2\mathbb{Z}\}^n$, $\boldsymbol{\pi}(\boldsymbol{\phi}_{\boldsymbol{j}}) = \boldsymbol{j}$ is a group isomorphism from $(G, \odot)$ to $(\{\mathbb{Z}/2\mathbb{Z}\}^n, +)$ because

$$\boldsymbol{\pi}(\boldsymbol{\phi}_{\boldsymbol{j}} + \boldsymbol{\phi}_{\boldsymbol{j}'}) = \boldsymbol{\pi}(\boldsymbol{\phi}_{\boldsymbol{j}+\boldsymbol{j}'}) = \boldsymbol{j} + \boldsymbol{j}' = \boldsymbol{\pi}(\boldsymbol{\phi}_{\boldsymbol{j}}) + \boldsymbol{\pi}(\boldsymbol{\phi}_{\boldsymbol{j}'}).$$

□

On the other hand, the Fourier matrix can be expressed as a Vandermonde matrix: $[\boldsymbol{\Phi}]_{j,k} = e^{-i2\pi(j\cdot k)/N}$, where $\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^*/N$ and $\boldsymbol{\Phi}^T = \boldsymbol{\Phi}^*$.

Proposition 2.2. *(modulation) Suppose $H = \{\boldsymbol{\phi}_j | \forall j\}$ is a set where $\boldsymbol{\phi}_j$ is the jth column of STFT matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N\times N}$. Then $(H, \odot)$ is an abelian group that is isomorphic to the integers modulo group $\mathbb{Z}/N\mathbb{Z}$ under addition.*

Proof. Hence forth, let the column index of STFT be represented via modulus $j \in \mathbb{Z}/N\mathbb{Z}$. From the definition of $\boldsymbol{\Phi}$,

$$\boldsymbol{\phi}_j \odot \boldsymbol{\phi}_j' = \begin{bmatrix} \vdots \\ e^{i2\pi(j+j')\cdot k/N} \\ \vdots \end{bmatrix} = \boldsymbol{\phi}_{j+j'}$$

As in the previous proof, this satisfies the axioms of abelian group, and the bijective map $\rho : H \to \mathbb{Z}/N\mathbb{Z}$, $\rho(\boldsymbol{\phi}_j) = j$ is a group isomorphism from $(H, \odot)$ to $(\mathbb{Z}/N\mathbb{Z}, +)$. □

Corollary 2.3. *(commutativity) Suppose $\boldsymbol{\Phi} \in \mathbb{R}^{N\times N}$ is a HFT or Fourier matrix, and $\boldsymbol{\phi}_{\boldsymbol{j}}$ is its $\boldsymbol{j}$th column as before. Then $\boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{\phi}_{\boldsymbol{j}}^*) = \boldsymbol{P}_{\boldsymbol{j}}\boldsymbol{\Phi}$, where $\boldsymbol{P}_{\boldsymbol{j}}$ is a permutation matrix whose $\boldsymbol{k}$th row is $\boldsymbol{e}_{\boldsymbol{j}+\boldsymbol{k}}^T$.*

Proof. Recall $\boldsymbol{\Phi}$ is symmetric. Then,

$$\boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{\phi}_{\boldsymbol{j}}^*) = \begin{bmatrix} \vdots \\ \boldsymbol{\phi}_{\boldsymbol{k}}^{*T} \\ \vdots \end{bmatrix} \mathrm{diag}(\boldsymbol{\phi}_{\boldsymbol{j}}^*) = \begin{bmatrix} \vdots \\ \{\boldsymbol{\phi}_{\boldsymbol{k}}^* \odot \boldsymbol{\phi}_{\boldsymbol{j}}^*\}^T \\ \vdots \end{bmatrix}.$$

From Proposition 2.1 or 2.2,

$$\boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{\phi}_{\boldsymbol{j}}^*) = \begin{bmatrix} \vdots \\ \boldsymbol{\phi}_{\boldsymbol{j}+\boldsymbol{k}}^{*T} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \boldsymbol{e}_{\boldsymbol{j}+\boldsymbol{k}}^T \\ \vdots \end{bmatrix} \boldsymbol{\Phi} = \boldsymbol{P}_{\boldsymbol{j}}\boldsymbol{\Phi}.$$

□

Note that $\boldsymbol{P}_{\boldsymbol{0}}$ is an identity matrix, and the above result naturally extends to discrete cosine and short-time Fourier transforms.

3. NOISE MODEL IN TRANSFORM DOMAIN

Recall (1) and, acknowledging the abuse of notation, let $\boldsymbol{h}(\boldsymbol{f})$ indicate an element-wise operation of $h(\cdot) : \mathbb{R} \to \mathbb{R}$ on $\boldsymbol{f}$. Then the vectorized version of (1) is $\boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}\big(\boldsymbol{f}, \mathrm{diag}(\boldsymbol{h}(\boldsymbol{f})^2)\big)$. Let $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{f}$ be the ideal transform coefficients, and $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{g}$ is the observed coefficients. Recall the inverse transform:

$$\boldsymbol{f} = (1/N)\boldsymbol{\Phi}^T\boldsymbol{x} = (1/N)\sum_{\boldsymbol{j}} \boldsymbol{\phi}_{\boldsymbol{j}}^* x_{\boldsymbol{j}}.$$

The classical interpretation is that $\boldsymbol{f}_a = (x_0/N)\boldsymbol{\phi}_0^*$ is the "approximation" of f, where x_0 is the lowest frequency (scaling) coefficient of $\boldsymbol{x}$ corresponding to $\boldsymbol{\phi}_0^* = (1, \cdots, 1)^T$. Assuming sufficient smoothness, we expand $\boldsymbol{h}(\boldsymbol{f})$ via the Taylor series about $\boldsymbol{f}_a$

$$\begin{aligned}
\boldsymbol{g} &= \boldsymbol{f} + \mathrm{diag}\Big(\boldsymbol{h}(\boldsymbol{f}_a) + \boldsymbol{h}'(\boldsymbol{f}_a)(\boldsymbol{f} - \boldsymbol{f}_a) + \cdots\Big)\boldsymbol{\varepsilon} \\
&= \boldsymbol{f} + \mathrm{diag}\Big(h(x_0/N)\boldsymbol{\phi}_0^* + h'(x_0/N)(\boldsymbol{f}_d) + \cdots\Big)\boldsymbol{\varepsilon},
\end{aligned}$$

where $h'(f) = dh/df$, etc., and $\boldsymbol{f}_d = \boldsymbol{f} - \boldsymbol{f}_a = (1/N)\sum_{\boldsymbol{j}\neq 0} \boldsymbol{\phi}_{\boldsymbol{j}}^* x_{\boldsymbol{j}}$ is often regarded as the "detail" of $\boldsymbol{f}$ because it is a linear combination of the higher frequency coefficients of $\boldsymbol{x}$ only. Using the Taylor series representation and the commutative property of the matrix $\boldsymbol{\Phi}$ in Corollary 2.3, the transform of the observed signal $\boldsymbol{g}$ is:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{\Phi}\boldsymbol{f} + \boldsymbol{\Phi}\,\mathrm{diag}\Big(h(x_0/N)\boldsymbol{\phi}_0^* + \frac{h'(x_0/N)}{N}\sum_{\boldsymbol{j}\neq 0} \boldsymbol{\phi}_{\boldsymbol{j}}^* x_{\boldsymbol{j}} + \cdots\Big)\boldsymbol{\varepsilon} \\
&= \boldsymbol{x} + \Big(h(x_0/N) + \frac{h'(x_0/N)}{N}\sum_{\boldsymbol{j}\neq 0} x_{\boldsymbol{j}}\boldsymbol{P}_{\boldsymbol{j}} + \cdots\Big)\boldsymbol{\Phi}\boldsymbol{\varepsilon} \\
&= \boldsymbol{x} + \boldsymbol{M}(\boldsymbol{x})\tilde{\boldsymbol{\varepsilon}}. \qquad (2)
\end{aligned}$$

Here $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\Phi}\boldsymbol{\varepsilon} \sim \mathcal{N}(0, N\boldsymbol{I})$, and $\boldsymbol{M}(\boldsymbol{x})$ is a polynomial matrix:

$$\boldsymbol{M}(\boldsymbol{x}) = \boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{f}))\boldsymbol{\Phi}^{-1} = \lim_{K\to\infty}\sum_{k=0}^{K}\frac{h^{(k)}(x_0/N)}{k!N^k}\left(\widetilde{\boldsymbol{M}}(\boldsymbol{x})\right)^k,$$

where $h^{(k)}(f) = d^k h(f)/df^k$ and $\widetilde{\boldsymbol{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{j}\neq 0} x_{\boldsymbol{j}}\boldsymbol{P}_{\boldsymbol{j}}$. When the Taylor series is carried out to infinity, (2) asymptotically characterizes the interaction between the signal and noise in (1) in the transform domain, and it is exact when $h^{(k)}(\cdot) = 0, \forall k > K$ for some K. In practice, $\boldsymbol{M}(\boldsymbol{x})$ is approximated to the Kth order polynomial. In fact, the existing practice of local AWGN modeling (as described in Section 1) is the zeroth order polynomial approximation of $\boldsymbol{M}(\boldsymbol{x})$; similarly, the Haar-Fisz algorithm is analogous to iteratively scaling $\boldsymbol{y}$ and $\boldsymbol{x}$ by $h(x_0/N)^{-1}$ [8,9]. However, the model in (2) suggests that we can achieve a higher degree of precision.

4. STATISTICAL ANALYSIS AND ESTIMATION

4.1. Likelihood Function and Maximum *A Posteriori* Estimator

We managed to largely bypass the nonlinear function $h(\cdot)$ in (2) because $\boldsymbol{y}$ is now a polynomial function of $\boldsymbol{x}$. This admits a direct manipulation of $\boldsymbol{y}$ based on the posterior distribution of the latent variable as deduced from the the likelihood of observed coefficients. That is, $\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}\big(\boldsymbol{x}, N\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T\big)$; and given a choice of the prior $p(\boldsymbol{x})$, the posterior distribution of $\boldsymbol{x}$ conditioned on $\boldsymbol{y}$ is:

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &= \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}\\
&= \frac{p(\boldsymbol{x})/p(\boldsymbol{y})\exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{x})^T(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T)^{-1}(\boldsymbol{y}-\boldsymbol{x})}{2N}\right)}{(2\pi)^{N/2}|N\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T|^{1/2}}\\
p(\boldsymbol{y}) &= \int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}\\
&= \int\frac{p(\boldsymbol{x})\exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{x})^T(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T)^{-1}(\boldsymbol{y}-\boldsymbol{x})}{2N}\right)d\boldsymbol{x}}{(2\pi)^{N/2}|N\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T|^{1/2}}.
\end{aligned}$$

This posterior probability distribution gives rise to new techniques for statistical inference and estimation. For example, the standard form of MAP estimator is:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}) = \arg\max_{\boldsymbol{x}}\log p(\boldsymbol{y}|\boldsymbol{x}) + \log p(\boldsymbol{x}) - \log p(\boldsymbol{y}).$$

Below, we sketch out a method to solve for the optimal $\boldsymbol{x}$ iteratively using the Levenberg-Marquardt algorithm [14–16]. Given the current estimate $\hat{\boldsymbol{x}}^{\text{old}}$ and a parameter λ, the next iterate estimate is defined in terms of the increment $\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{\text{new}} - \hat{\boldsymbol{x}}^{\text{old}}$, where

$$\tilde{\boldsymbol{x}} = (\boldsymbol{J}^T\boldsymbol{J} + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}^T\Big[\log p(\boldsymbol{y}|\hat{\boldsymbol{x}}^{\text{old}}) + \log p(\hat{\boldsymbol{x}}^{\text{old}})\Big].$$

Here, $\boldsymbol{J}$ is the Jacobian of $\log p(\boldsymbol{x}|\boldsymbol{y})$ evaluated at $\hat{\boldsymbol{x}}^{\text{old}}$:

$$\begin{aligned}
\boldsymbol{J} &= \left[\tfrac{d\log p(\boldsymbol{x}|\boldsymbol{y})}{dx_0}\quad\cdots\quad\tfrac{d\log p(\boldsymbol{x}|\boldsymbol{y})}{dx_k}\quad\cdots\right]\Big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}^{\text{old}}}\\
\frac{d\log p(\boldsymbol{x}|\boldsymbol{y})}{dx_{\boldsymbol{j}}} &= \frac{d\log p(\boldsymbol{y}|\boldsymbol{x})}{dx_{\boldsymbol{j}}} + \frac{d\log p(\boldsymbol{x})}{dx_{\boldsymbol{j}}}.
\end{aligned}$$

The derivative $\frac{d\log p(\boldsymbol{y}|\boldsymbol{x})}{dx_{\boldsymbol{j}}}$ is the sum of the following two quantities:

$$\begin{aligned}
&\frac{d\log|N\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T|}{dx_{\boldsymbol{j}}} = -2\mathrm{Tr}\left[\boldsymbol{M}(\boldsymbol{x})^{-1}\frac{d\boldsymbol{M}(\boldsymbol{x})}{dx_{\boldsymbol{j}}}\right]\\
&\frac{d(\boldsymbol{y}-\boldsymbol{x})^T(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T)^{-1}(\boldsymbol{y}-\boldsymbol{x})}{dx_{\boldsymbol{j}}}\\
&\quad= 2\boldsymbol{e}_{\boldsymbol{j}}^T(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T)^{-1}(\boldsymbol{y}-\boldsymbol{x})\\
&\qquad- 2(\boldsymbol{y}-\boldsymbol{x})^T\boldsymbol{M}(\boldsymbol{x})^{-1}\frac{d\boldsymbol{M}(\boldsymbol{x})^T}{dx_{\boldsymbol{j}}}(\boldsymbol{M}(\boldsymbol{x})\boldsymbol{M}(\boldsymbol{x})^T)^{-1}(\boldsymbol{y}-\boldsymbol{x}),
\end{aligned}$$

where

$$\frac{d\boldsymbol{M}(\boldsymbol{x})}{dx_{\boldsymbol{j}}} = \frac{h'(x_0/N)}{N}\boldsymbol{P}_{\boldsymbol{j}} + \frac{h''(x_0/N)}{N^2}\sum_{\boldsymbol{j}'\neq 0}x_{\boldsymbol{j}'}\boldsymbol{P}_{\boldsymbol{j}+\boldsymbol{j}'} + \cdots.$$

Combined with the first derivative of the prior $p(\boldsymbol{x})$, the implementation of this MAP estimator is straightforward.

4.2. Optimal Mean Estimator Selection for Poisson Process

Owing partly to Stein's contributions to estimation theories, parametric estimators are widely popular [17–20]. The celebrated result of Stein's unbiased risk estimate (SURE) for a vector $\boldsymbol{f}$ states that given a noisy observation $\boldsymbol{g}|\boldsymbol{f} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(f_m, 1)$, the expected squared error of a parametric estimator $\hat{\boldsymbol{f}}_\theta(\boldsymbol{g}) = \boldsymbol{g} + \boldsymbol{\psi}_\theta(\boldsymbol{g})$ is

$$E\Big[\|\boldsymbol{f} - \hat{\boldsymbol{f}}_\theta(\boldsymbol{g})\|^2\Big|\boldsymbol{f}\Big] = N + E\Big[\|\boldsymbol{\psi}_\theta(\boldsymbol{g})\|^2 + 2\nabla_g\cdot\boldsymbol{\psi}_\theta(\boldsymbol{g})\Big|\boldsymbol{f}\Big],$$

where $\nabla_g\cdot\boldsymbol{\psi}_\theta(\boldsymbol{g}) = \sum_i\frac{d}{dg_i}\psi_{\theta,i}$ and $\boldsymbol{\psi}_\theta = (\psi_{\theta,0},\cdots,\psi_{\theta,N})^T$. The optimal choice of parameter θ in $\boldsymbol{\psi}_\theta$ is therefore the minimizer of this function. The obvious advantage to this approach is that θ can be adapted to $\boldsymbol{f}$ even when its smoothness is unknown [18]. A more general form is developed by Raphan *et al.* [21]

We work with (2) to derive a variant to the SURE quantity above that is amenable to estimating the mean of the Poisson process in the HFT and Fourier domains [22,23].

Proposition 4.1. *Suppose $\boldsymbol{\Phi}$ is the HFT or Fourier matrix, $h(f_m) = f_m^{1/2}$, and $\boldsymbol{M}(\boldsymbol{x}) = \boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{\Phi}^{-1}\boldsymbol{x}))\boldsymbol{\Phi}^{-1}$. If $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{M}(\boldsymbol{x})\tilde{\boldsymbol{\epsilon}}$ then $E[\boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{M}(\boldsymbol{x})] = E[\boldsymbol{M}(\boldsymbol{y})^T\boldsymbol{M}(\boldsymbol{y})]$.*

Proof. $E[\boldsymbol{M}(\boldsymbol{y})^T\boldsymbol{M}(\boldsymbol{y})]$

$$\begin{aligned}
&= E[\boldsymbol{\Phi}^{\text{-}T}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{\Phi}^{\text{-}1}\boldsymbol{y}))^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{\Phi}^{\text{-}1}\boldsymbol{y}))\boldsymbol{\Phi}^{\text{-}1}]\\
&= NE[\boldsymbol{\Phi}^{\text{-}T}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{g}))^2\boldsymbol{\Phi}^{\text{-}1}] = N\boldsymbol{\Phi}^{\text{-}T}\,\mathrm{diag}(E[\boldsymbol{f} + \boldsymbol{h}(\boldsymbol{f})\boldsymbol{\epsilon}])\boldsymbol{\Phi}^{\text{-}1}\\
&= E[\boldsymbol{\Phi}^{\text{-}T}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{f}))^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{h}(\boldsymbol{f}))\boldsymbol{\Phi}^{\text{-}1}]\\
&= E[\boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{M}(\boldsymbol{x})].
\end{aligned}$$

□

Corollary 4.2. *(modified SURE) Suppose $g_m|\boldsymbol{f} \overset{i.i.d.}{\sim} \mathcal{P}(f_m)$, and $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{f}$, $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{g}$. Let $\hat{\boldsymbol{x}}_\theta(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{\psi}_\theta(\boldsymbol{y})$ be a weakly differentiable parametric estimator of $\boldsymbol{x}$ such that $d\boldsymbol{\psi}_\theta(\boldsymbol{y})/d\boldsymbol{y}$ is piece-wise constant. Then the expected risk is:*

$$\begin{aligned}
E\Big[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_\theta(\boldsymbol{y})\|^2\Big|\boldsymbol{x}\Big] &= E\Big[\boldsymbol{\psi}_\theta(\boldsymbol{y})^T\boldsymbol{\psi}_\theta(\boldsymbol{y})\Big|\boldsymbol{x}\Big]\\
&\quad+ Tr\left(E\Big[\boldsymbol{M}(\boldsymbol{y})^T\boldsymbol{M}(\boldsymbol{y})\Big|\boldsymbol{x}\Big]\left(N\boldsymbol{I} + \frac{2}{N}\frac{d\boldsymbol{\psi}_\theta(\boldsymbol{y})}{d\boldsymbol{y}}\right)\right),
\end{aligned}$$

where $\boldsymbol{M}(\boldsymbol{x})$ is as defined in (2).

Proof. The mean squared error can be rewritten as:

$$E\Big[\|\boldsymbol{x}-\hat{\boldsymbol{x}}_\theta(\boldsymbol{y})\|^2\Big|\boldsymbol{x}\Big] = E\Big[\|\boldsymbol{M}(\boldsymbol{x})\tilde{\varepsilon}+\boldsymbol{\psi}_\theta(\boldsymbol{y})\|^2\Big|\boldsymbol{x}\Big]$$
$$= E\Big[\tilde{\varepsilon}^T\boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{M}(\boldsymbol{x})\tilde{\varepsilon}+2\tilde{\varepsilon}^T\boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{\psi}_\theta(\boldsymbol{y})+\boldsymbol{\psi}_\theta(\boldsymbol{y})^T\boldsymbol{\psi}_\theta(\boldsymbol{y})\Big|\boldsymbol{x}\Big].$$

By Proposition 4.1, the first term equals to $N\mathrm{Tr}(E[\boldsymbol{M}(\boldsymbol{y})^T\boldsymbol{M}(\boldsymbol{y})|\boldsymbol{x}])$. To reconcile the second term, we borrow the following from Stein:

$$E\Big[\tilde{\varepsilon}^T\boldsymbol{\eta}_\theta(\tilde{\varepsilon})\Big|\boldsymbol{x}\Big] = \frac{1}{N}E\Big[\nabla_{\tilde{\varepsilon}}\cdot\boldsymbol{\eta}_\theta(\tilde{\varepsilon})\Big|\boldsymbol{x}\Big],$$

where $\boldsymbol{\eta}_\theta:\mathbb{R}^N\to\mathbb{R}^N$ is a weakly differentiable function. Recall $d\boldsymbol{\psi}_\theta(\boldsymbol{y})/d\boldsymbol{y}$ is a constant, and suppose we set $\boldsymbol{\eta}_\theta(\tilde{\varepsilon}) = \boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{\psi}_\theta(\boldsymbol{x}+\boldsymbol{M}(\boldsymbol{x})\tilde{\varepsilon})$. Then using chain rule:

$$E\Big[\tilde{\varepsilon}^T\boldsymbol{M}(\boldsymbol{x})^T\boldsymbol{\psi}_\theta(\boldsymbol{y})\Big|\boldsymbol{x}\Big] = \frac{1}{N}E\Big[\sum_j\frac{d}{d\tilde{\varepsilon}_j}\eta_{\theta j}(\tilde{\varepsilon})\Big|\boldsymbol{x}\Big]$$
$$= \frac{1}{N}E\Big[\sum_j\left(\frac{d\eta_{\theta j}(\tilde{\varepsilon})}{d\boldsymbol{\psi}_\theta(\boldsymbol{y})}\right)\left(\frac{d\boldsymbol{\psi}_\theta(\boldsymbol{y})}{d\boldsymbol{y}}\right)\left(\frac{d\boldsymbol{y}}{d\tilde{\varepsilon}_j}\right)\Big|\boldsymbol{x}\Big]$$
$$= \frac{1}{N}E\Big[\sum_j\boldsymbol{e}_j^T\boldsymbol{M}(\boldsymbol{x})^T\left(\frac{d\boldsymbol{\psi}_\theta(\boldsymbol{y})}{d\boldsymbol{y}}\right)\boldsymbol{M}(\boldsymbol{x})\boldsymbol{e}_j\Big|\boldsymbol{x}\Big]$$
$$= \frac{1}{N}\mathrm{Tr}\Big(E\left[\boldsymbol{M}(\boldsymbol{y})^T\boldsymbol{M}(\boldsymbol{y})\Big|\boldsymbol{x}\right]\frac{d\boldsymbol{\psi}_\theta(\boldsymbol{y})}{d\boldsymbol{y}}\Big),$$

where the properties of matrix trace and the results of Proposition 4.1 are used in the last step. □

5. CONCLUSION

This paper addressed the transform domain modeling of signal-dependent noise. The noise-corrupted signal has an asymptotic representation in the transform domain when $h(\cdot)$ is a K-times differentiable function. The posterior distribution of these filterbank or Fourier coefficients is readily accessible because both the prior and the likelihood are encoded in the transform domain, and the maximum *a posteriori* estimator naturally stems from such analysis. The structure of noise in the transform domain admits a variant of Stein's unbiased estimate of risk conducive to transform-based processing of a signal corrupted by signal-dependent noise.

6. REFERENCES

[1] K. Hirakawa and X.-L. Meng, "An empirical Bayes EM-wavelet unification for simultaneous denoising, interpolation, and/or demosaicing," *IEEE Int'l Conf Image Processing*, 2006.

[2] J. Portilla, D. Otaduy, and C. Dorronsoro, "Low-complexity linear demosaicing using joint spatial-chromatic image statistics," *IEEE Int'l Conf on Image Processing*, September 2005.

[3] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Int'l Conf on Acoustics, Speech, and Signal Processing*, 1998.

[4] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.

[5] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Processing*, vol. 50, 2002.

[6] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, 4th edition, 2002.

[7] A. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing*, Prentice-Hall, Inc., 1989.

[8] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Comp. Graph. Stat.*, vol. 13, pp. 621–638, 2004.

[9] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloquium Mathematicum*, vol. 3, pp. 138–146, 1955.

[10] C. Kervrann and A. Trubuil, "An adaptive window approach for poisson noise reduction and structure preserving in confocal microscopy," in *Biomedical Imaging: Macro to Nano, 2004. IEEE Int'l Symposium on*, 2004, vol. 1, pp. 788–791.

[11] C. Stein, "Confidence sets for the mean of a multivariate normal distribution (with discussion)," *J. Roy. Statist. Soc. Ser. B*, vol. 24, pp. 265–296, 1962.

[12] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. Third Berlzeley Symp. Math. Statist. Probab.*, 1956, vol. 1, pp. 197–206.

[13] C. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.

[14] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.

[15] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.

[16] P. E. Gill and W. Murray, "Algorithms for the solution of the nonlinear least-squares problem," *SIAM J. Numer. Anal.*, vol. 15, no. 5, pp. 977–992, 1978.

[17] I. M. Johnstone, "On minimax estimation of a sparse normal mean vector," *The Annals of Statistics*, vol. 22, no. 1, pp. 271–289, 1994.

[18] David L. Donoho and Iain M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[19] Y.-C. Chang, B.-K. Shyu, C.-Y. Cho, and J.-S. Wang, "Adaptive post-processing for region-based fractal image compression," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[20] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society B*, vol. 59, no. 2, pp. 319–351, 1997.

[21] M. Raphan and E. P. Simoncelli, "Learning to be bayesian without supervision," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds., pp. 1145–1152. MIT Press, Cambridge, MA, 2007.

[22] X.-L. Meng, "From unit root to Stein's estimator to Fisher's k statistics: If you have a moment, I can tell you more," *Statistical Science*, vol. 20, no. 2, pp. 141–162, 2005.

[23] S. M. Stigler, "The 1988 neyman memorial lecture: A galtonian perspective on shrinkage estimators," *Statistical Science*, vol. 5, pp. 147–155, 1990.