SUPLEMENTARY MATERIAL FOR
DEEP UNSUPERVISED DESPECKLING WITH UNBIASED RISK ESTIMATION

1. NOTATIONS

Throughout the paper and the supplementary material, we use the following notations:

1. $\mathbf{Y}$ to denote m dimensional multivariate (vector) random variables.
2. $\mathbf{y}$ to denote m dimensional samples from the corresponding random variables.
3. Y to denote scalar random variables.
4. y to denote samples from the corresponding scalar random variable or a constant depending on the context.
5. Y_i is i^{th} scalar random variable of a vector random variable $\mathbf{Y}$.
6. y_i is i^{th} pixel represented as realization of a scalar random variable Y.
7. $\mathbb{E}$ is expectation operator (underlying random variable is clear from context or mentioned explicitly), $\odot$ is element-wise (Hadamard) product.

2. DETAILS ON UNBIASED RISK ESTIMATION

2.1. Oracle MSE

Our aim is to obtain an estimate of $\mathbf{x}$ (a realization of $\mathbf{X}$), given the measurement $\mathbf{y}$, which is a realization of random image variable $\mathbf{Y}$. We denote this estimate as a function of observables, $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{Y})$. In general, $\mathbf{f}$ may be any linear or non-linear, parametric or non-parametric function. The criterion which we choose to minimize is the ensemble-averaged mean-square error (or *risk*) between $\mathbf{x}$ and $\hat{\mathbf{x}}$

$$\zeta(\mathbf{f}) = \frac{1}{m}\mathbb{E}\{\|\mathbf{f}(\mathbf{Y}) - \mathbf{x}\|^2\} = \frac{1}{m}\sum_{i=1}^{m}(x_i - f_i(\mathbf{Y}))^2, \tag{1}$$

which requires knowledge of the ground-truth reflectance $\mathbf{x}$. Consider the expansion of Eq. (1) (without the factor $1/m$),

$$\zeta(\mathbf{f}) = \|\mathbf{x}\|^2 + \mathbb{E}\{\|\mathbf{f}(\mathbf{Y})\|^2\} - 2\sum_{i=1}^{m}\mathbb{E}\{x_i f_i(\mathbf{Y})\}, \tag{2}$$

where $f_i(\mathbf{Y})$ denotes the i^{th} entry of the denoised image. Since the optimization is carried out with respect to $\mathbf{f}$, the deterministic (but unknown) factor $\|\mathbf{x}\|^2$ does not play a role (in contrast with the Bayesian framework, where a prior is assumed on $\mathbf{x}$). On the other hand, the term $\mathbb{E}\{x_i f_i(\mathbf{Y})\}$ depends on the unknowns x_i and hence a direct optimization is not possible without knowledge of the ground truth image. Throughout our work, we call this version of the cost the *Oracle MSE estimate*.

2.2. Proof of Corollary 1.1 (our result from main paper)

To estimate the Oracle MSE without the ground truth, unbiased risk estimation methods can be applied. Seelamantula and Blu [1] first presented a surrogate risk for the case of multiplicative Gamma distributed noise model called Multiplicative Unbiased Risk Estimate (MURE). In this section, We present a detailed proof of corollary 1.1 from the main paper.

Theorem 1. *(Multivariate version) Let $\mathbf{Y} = \mathbf{x}\mathbf{N}$, where $\mathbf{x} \in \mathbb{R}_+^{\mathbf{m}}$ is deterministic but unknown reflectance image. Let $\mathbf{Y}, \mathbf{N} \in \mathbb{R}_+^m$, and $\mathbf{N} \sim \Gamma(k, k)$ with independent entries, then, the vector random variable*

$$\hat{\zeta}(\mathbf{f}) = \frac{k}{k+1}\|\mathbf{Y}\|^2 - 2\mathbf{Y}^{\mathrm{T}}\mathcal{M}\mathbf{f}(\mathbf{Y}) + \|\mathbf{f}(\mathbf{Y})\|^2 \tag{3}$$

is an unbiased estimator of the MSE, $\zeta(\mathbf{f}) = \mathbb{E}_{\mathbf{N}}\{\|\mathbf{f}(\mathbf{Y}) - \mathbf{x}\|^{\mathbf{2}}\}$, where $\mathbb{E}$ is the expectation operator. For a scalar function $f(Y)$, the operator $\mathcal{M}$ is defined as $\mathcal{M}f(Y) = k\int_0^1 s^{k-1} f(sY)ds$. This notation is extended straightforwardly to multivariate vector functions $\mathbf{f}(\mathbf{Y}) = [f_1(\mathbf{Y}), f_2(\mathbf{Y}), \ldots, f_m(\mathbf{Y})]^{\mathrm{T}}$ according to $\mathcal{M}\mathbf{f}(\mathbf{Y}) = [\mathcal{M}_1 f_1(\mathbf{Y}), \mathcal{M}_2 f_2(\mathbf{Y}), \ldots, \mathcal{M}_m f_m(\mathbf{Y})]^{\mathrm{T}}$, where $\mathcal{M}_i \mathbf{f_i}(\mathbf{Y})$ applies the operator $\mathcal{M}$ to the i^{th} input component of $\mathbf{f}(\mathbf{Y})$ only.

Corollary 1.1. *(Multivariate, series version of MURE) Let $\mathbf{Y} = \mathbf{x}\mathbf{N}$, where $\mathbf{x} \in \mathbb{R}_+^{\mathbf{m}}$ is deterministic but unknown reflectance image. Let $\mathbf{Y}, \mathbf{N} \in \mathbb{R}_+^m$, and $\mathbf{N} \sim \Gamma(k, k)$ with independent entries, then, the vector random variable*

$$\hat{\zeta}(\mathbf{f}) = \frac{k}{k+1}\|\mathbf{Y}\|^2 + \|\mathbf{f}(\mathbf{Y})\|^2 - 2\sum_{i=1}^{m}\sum_{p=0}^{\infty}(-1)^p \frac{k!}{(k+p)!} Y_i^{p+1} \frac{\partial \mathbf{f}_i^{(p)}(\mathbf{Y})}{\partial Y_i}, \tag{4}$$

is an unbiased estimator of the MSE, $\zeta(\mathbf{f}) = \mathbb{E}_{\mathbf{N}}\{\|\mathbf{f}(\mathbf{Y}) - \mathbf{x}\|^{\mathbf{2}}\}$, where $\mathbb{E}$ is the expectation operator. p is the order of partial derivative of $\mathbf{f}_i(\mathbf{Y})$ w.r.t. $\mathbf{Y}_i$, the i^{th} pixel of the input noisy image.

Proof. We first prove the case for a scalar function $f : \mathbb{R} \to \mathbb{R}$. We have:

$$\mathbb{E}\{(f(Y) - x)^2\} \quad = \quad \mathbb{E}\{f(Y)^2\} + \mathbb{E}\{x^2\} - 2\mathbb{E}\{xf(Y)\}.$$

Expanding term by term,

- $\mathbb{E}\{Y^2\} = \mathbb{E}\{x^2 n^2\} = \frac{k+1}{k}\mathbb{E}\{x^2\}$, and hence

$$\mathbb{E}\{x^2\} = \frac{k}{k+1}\mathbb{E}\{Y^2\}. \tag{5}$$

- We have, $\mathbb{E}\{xf(Y)\}$

$$\begin{aligned}
&= \int_{0^+}^{\infty} x\ f(Y)\ f_N(n)dn \\
&= \int_{0^+}^{\infty} x\ f(Y)\ \frac{k^k}{\Gamma(k)} n^{k-1} e^{-kn} dn \\
&= \int_{0^+}^{\infty} kx\ f(Y)\ \frac{n^k k^k}{\Gamma(k)} \left[\frac{e^{-kn}}{nk}\right] dn \\
&= \int_{0^+}^{\infty} kx\ f(Y)\ \left[\int_{0^+}^{1} \frac{n^k k^k}{\Gamma(k)}\ \frac{1}{s^2} e^{-kn/s} ds\right] dn \\
&= \int_{0^+}^{\infty} kx\ \left[\int_{0^+}^{1} f(xn)\ \frac{ns^{k-1}}{s^2} f_N(n/s) ds\right] dn,
\end{aligned}$$

changing the order of integration and substituting $p = n/s$,

$$
\begin{aligned}
& \int_{0^+}^{\infty} kx \left[\int_{0^+}^{1} f(xn) \frac{ns^{k-1}}{s^2} f_N(n/s) ds \right] dn \\
= & \int_{0^+}^{1} \int_{0^+}^{\infty} kxp\, f(xps)\, s^{k-1} f_N(p)\, dp\, ds \\
= & \int_{0^+}^{\infty} Y \left[k \int_{0^+}^{1} f(sY)\, s^{k-1} ds \right] f_N(p) dp, \\
= & \mathbb{E}\{Y \mathcal{M} f(Y)\}. \qquad (6)
\end{aligned}
$$

where we applied change of order of integration (assuming conditions of Fubini's theorem to be true and that the limit exists at 0) again, and substituted $Y = xp$. Using (5) and (6), we have, $\mathbb{E}\{(f(Y) - x)^2\}$

$$
\begin{aligned}
= & \mathbb{E}\left\{ f(Y)^2 - 2Y\mathcal{M}f(Y) + \frac{k}{k+1} Y^2 \right\}. \\
= & \mathbb{E}\{\hat{\zeta}(f)\}.
\end{aligned}
$$

Which shows that the MURE cost $\hat{\zeta}(f)$ is an unbiased estimator of the oracle cost $\zeta(f)$. For a scalar function $f(Y)$, the operator $\mathcal{M}$ is defined as $\mathcal{M}f(Y) = k\int_0^1 s^{k-1} f(sY) ds$. This notation is extended straightforwardly to multivariate vector functions $\mathbf{f}(\mathbf{Y}) = [f_1(\mathbf{Y}), f_2(\mathbf{Y}), \ldots, f_m(\mathbf{Y})]^{\mathrm{T}}$ according to $\mathcal{M}\mathbf{f}(\mathbf{Y}) = [\mathcal{M}_1 f_1(\mathbf{Y}), \mathcal{M}_2 f_2(\mathbf{Y}), \ldots, \mathcal{M}_m f_m(\mathbf{Y})]^{\mathrm{T}}$, where $\mathcal{M}_i \mathbf{f_i}(\mathbf{Y})$ applies the operator $\mathcal{M}$ to the i^{th} input component of $\mathbf{f}(\mathbf{Y})$ only. Hence, the multivariate result is straightforward to obtain by applying the scalar version of the estimator in Theorem 1 of the main paper to the individual components of the cost function in (1). Thus the cost for a vector function can be written as

$$
\hat{\zeta}(\mathbf{f}) = \frac{k}{k+1}\|\mathbf{Y}\|^2 + \|\mathbf{f}(\mathbf{Y})\|^2 - 2\mathbf{Y}^{\mathrm{T}}\mathcal{M}\mathbf{f}(\mathbf{Y}) \qquad (7)
$$

For the series approximation, we note that the cross term operator $\mathcal{M}$ for a vector to vector function can be expanded by applying the integration-by-parts operation:

$$
\begin{aligned}
\mathcal{M}_i f_i(\mathbf{Y}) &= \mathcal{M}_i f_i(Y_1, Y_2, ..., Y_i, ..., Y_m) \\
&= k\int_0^1 s^{k-1} f_i(Y_1, Y_2, ..., sY_i, ..., Y_m)\, \mathrm{d}s, \\
&= k\int_0^1 s^{k-1} f_i(\mathbf{S}_i\mathbf{Y})\, \mathrm{d}s, \\
&= s^k f_i(\mathbf{S}_i\mathbf{Y})\Big|_0^1 - \int_0^1 Y_i f_i^{'}(\mathbf{S}_i\mathbf{Y})\, s^k\, \mathrm{d}s, \\
&= f_i(Y_i) - \int_0^1 Y_i f_i^{'}(\mathbf{S}_i\mathbf{Y})\, s^{k+1}\, \mathrm{d}s, \\
&= f_i(\mathbf{Y}) - \left(\frac{s^{k+1}}{k+1} Y_i f_i^{'}(\mathbf{S}_i\mathbf{Y}) \Big|_0^1 - \int_0^1 Y_i f_i^{'}(\mathbf{S}_i\mathbf{Y})\, s^{k+1}\, \mathrm{d}s \right), \\
&= f_i(\mathbf{Y}) - \frac{1}{k+1} Y_i f_i^{'}(\mathbf{Y}) + \frac{1}{k+1}\int_0^1 Y_i f_i^{'}(\mathbf{S}_i\mathbf{Y})\, s^{k+1}\, \mathrm{d}s), \\
&= \sum_{p=0}^{\infty} (-1)^p \frac{k!}{(k+p)!} Y_i^p \frac{\partial f_i^{(p)}(\mathbf{Y})}{\partial Y_i}, \qquad (8)
\end{aligned}
$$

where $\mathbf{S}_i$ is a matrix constructed by replacing i^{th} diagonal element of identity matrix by s. Finally, writing all terms together for the vector version, we have:

$$\begin{aligned}
\mathbb{E}_{\mathbf{N}}\{\hat{\zeta}(\mathbf{f})\} &= \mathbb{E}_{\mathbf{N}}\left\{\frac{k}{k+1}\|\mathbf{Y}\|^2 + \|\mathbf{f}(\mathbf{Y})\|^2 - 2\sum_{i=1}^{m}\sum_{p=0}^{\infty}(-1)^p\frac{k!}{(k+p)!}Y_i^{p+1}\frac{\partial \mathbf{f}_i^{(p)}(\mathbf{Y})}{\partial Y_i}\right\} \\
&= \mathbb{E}_{\mathbf{N}}\left\{\frac{k}{k+1}\|\mathbf{Y}\|^2 + \|\mathbf{f}(\mathbf{Y})\|^2 - 2\mathbf{Y}^{\mathrm{T}}\mathcal{M}\mathbf{f}(\mathbf{Y})\right\} \\
&= \mathbb{E}_{\mathbf{N}}\left\{\|\mathbf{x}\|^2 + \|\mathbf{f}(\mathbf{Y})\|^2 - 2\sum_{i=1}^{m}x_i f_i(\mathbf{Y})\right\} \\
&= \mathbb{E}_{\mathbf{N}}\{\zeta(\mathbf{f})\}.
\end{aligned}$$

Thus, our proposed series-version of MURE is an unbiased estimator of the Oracle cost in Eq. 1. This completes the proof. □

3. MONTE CARLO ESTIMATION OF SERIES VERSION OF MURE

In this section, we state and prove the key results for Monte Carlo estimation of the MURE cost from the main paper.

Theorem 2. *Let* $\mathbf{f}(\mathbf{Y}) : \mathbb{R}^{\mathbf{m}} \to \mathbb{R}^{\mathbf{m}}$ *be a differentiable function and* $\mathbf{J_f}$ *denote its Jacobian. Let* $\mathbf{B} \in \mathbb{R}^{\mathbf{m}}$ *with i.i.d entries* $\sim \mathcal{N}(0,1)$. *Then,* $\sum_{i=1}^{m}[\mathbf{J_f}(\mathbf{Y})]_{ii}Y_i^2 = \lim_{\epsilon\to 0}\mathbb{E}_B\left\{(\mathbf{Y}\odot\mathbf{B})^{\mathbf{T}}\left(\frac{\mathbf{f}(\mathbf{Y}+\epsilon(\mathbf{Y}\odot\mathbf{B}))-\mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B}))}{\mathbf{2}\epsilon}\right)\right\}$.

Proof. Let $\mathbf{B} \in \mathbb{R}^{\mathbf{m}}$ with i.i.d entries $\sim \mathcal{N}(0,1)$. Applying the Taylor-series expansion, we have,

$$\mathbf{f}(\mathbf{Y}+\epsilon(\mathbf{Y}\odot\mathbf{B})) = \mathbf{f}(\mathbf{Y}) + \epsilon\,\mathbf{J_f}(\mathbf{Y})(\mathbf{Y}\odot\mathbf{B}) + \epsilon^2 h.o.t.\,, \tag{9}$$

A similar expansion exists for $\mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B}))$:

$$\mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B})) = \mathbf{f}(\mathbf{Y}) - \epsilon\,\mathbf{J_f}(\mathbf{Y})(\mathbf{Y}\odot\mathbf{B}) + \epsilon^2 h.o.t.\,, \tag{10}$$

Subtracting Eq. 10 from Eq. 9 and dividing by 2ϵ gives

$$\frac{\mathbf{f}(\mathbf{Y}+\epsilon(\mathbf{Y}\odot\mathbf{B})) - \mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B}))}{2\epsilon} = \mathbf{J_f}(\mathbf{Y})(\mathbf{Y}\odot\mathbf{B}) + \mathcal{O}(\epsilon^2).$$

Taking inner product with $(\mathbf{Y}\odot\mathbf{B})$ and computing the expectation w.r.t. $\mathbf{B}$ yields

$$\begin{aligned}
&\mathbb{E}_B\left\{(\mathbf{Y}\odot\mathbf{B})^{\mathbf{T}}\left(\frac{\mathbf{f}(\mathbf{Y}+\epsilon(\mathbf{Y}\odot\mathbf{B})) - \mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B}))}{\mathbf{2}\epsilon}\right)\right\} \\
&= \mathbb{E}_B\left\{(\mathbf{Y}\odot\mathbf{B})^{\mathbf{T}}\mathbf{J_f}(\mathbf{Y})(\mathbf{Y}\odot\mathbf{B})\right\} + \mathcal{O}(\epsilon^2) \\
&= \mathbb{E}_B\left\{\sum_{i=1}^{m}[\mathbf{J_f}(\mathbf{Y})]_{ii}Y_i^2 {b_i}^2 + \mathcal{O}(\epsilon^2)\right\} \\
&= \sum_{i=1}^{m}[\mathbf{J_f}(\mathbf{Y})]_{ii}Y_i^2 + \mathcal{O}(\epsilon^2)
\end{aligned}$$

Where the last step follows from the fact that $\mathbb{E}\{b_i b_j\} = 0$ and $\mathbb{E}\{b_i^2\} = 1$. Taking limit of this expression with $\epsilon \to 0$ yields the desired result. □

To actually compute this expectation, we utilize K realizations of $\mathbf{B}$ and n variations of ϵ. Hence, the estimate is evaluated as:

$$\sum_{i=1}^{m}[\mathbf{J_f}(\mathbf{Y})]_{ii}Y_i^2 = \frac{1}{nK}\sum_{r=1}^{K}\sum_{s=1}^{n}\left\{(\mathbf{Y}\odot\mathbf{B_r})^{\mathbf{T}}\left(\frac{\mathbf{f}(\mathbf{Y}+\epsilon_{\mathbf{s}}(\mathbf{Y}\odot\mathbf{B_r})) - \mathbf{f}(\mathbf{Y}-\epsilon_{\mathbf{s}}(\mathbf{Y}\odot\mathbf{B_r}))}{\mathbf{2}\epsilon_{\mathbf{s}}}\right)\right\}. \tag{11}$$

For all our experiments $n = K = 1$ yields reasonably good approximation with low computational complexity.

Theorem 3. *Let* $\mathbf{f}(\mathbf{Y}) : \mathbb{R}^{\mathbf{m}} \rightarrow \mathbb{R}^{\mathbf{m}}$ *be at a twice differentiable function and* $[\mathbf{H}_{\mathbf{f}}]_{\mathbf{i}}$ *denote the Hessian of* $\mathbf{f}_i$*. Let* $\mathbf{B} \in \mathbb{R}^{\mathbf{m}}$ *with i.i.d entries* $\sim Triangular(-2, 1)$*. Then* $\sum_{i=1}^{m} [\mathbf{H}_{\mathbf{f}}(\mathbf{Y})]_{iii} Y_i^3 = \lim_{\epsilon \to 0} \mathbb{E}_B \left\{ -5(\mathbf{Y} \odot \mathbf{B})^{\mathbf{T}} \left(\frac{\mathbf{f}(\mathbf{Y}+\epsilon(\mathbf{Y}\odot\mathbf{B}))-\mathbf{2f}(\mathbf{Y})+\mathbf{f}(\mathbf{Y}-\epsilon(\mathbf{Y}\odot\mathbf{B}))}{\epsilon^{\mathbf{2}}} \right) \right\}$.

Proof. Let $\mathbf{B} \in \mathbb{R}^{\mathbf{m}}$ with i.i.d entries $\sim Triangular(-2, 1)$. $Triangular(-2, 1)$ denotes the Triangular distribution with parameters $a = -2$ and $b = 1$ with moments $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = \frac{-1}{5}$. We have,

$$\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) = \mathbf{f}(\mathbf{Y}) + \epsilon\, \mathbf{J}_{\mathbf{f}}(\mathbf{Y})(\mathbf{Y} \odot \mathbf{B}) + \frac{1}{2}\epsilon^2 \begin{bmatrix} (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_1 (\mathbf{Y} \odot \mathbf{B}) \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_2 (\mathbf{Y} \odot \mathbf{B}) \\ \vdots \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_m (\mathbf{Y} \odot \mathbf{B}) \end{bmatrix} + \epsilon^3 h.o.t.\,. \tag{12}$$

Where $\mathbf{H}_i$ denotes hessian of the function $\mathbf{f}_i$. A similar expansion exists for $\mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B}))$:

$$\mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B})) = \mathbf{f}(\mathbf{Y}) - \epsilon\, \mathbf{J}_{\mathbf{f}}(\mathbf{Y})(\mathbf{Y} \odot \mathbf{B}) + \frac{1}{2}\epsilon^2 \begin{bmatrix} (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_1 (\mathbf{Y} \odot \mathbf{B}) \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_2 (\mathbf{Y} \odot \mathbf{B}) \\ \vdots \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_m (\mathbf{Y} \odot \mathbf{B}) \end{bmatrix} + \epsilon^3 h.o.t.\,. \tag{13}$$

Adding Eq. 12 and Eq. 13 gives:

$$\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) + \mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B})) = 2\, \mathbf{f}(\mathbf{Y}) + \epsilon^2 \begin{bmatrix} (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_1 (\mathbf{Y} \odot \mathbf{B}) \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_2 (\mathbf{Y} \odot \mathbf{B}) \\ \vdots \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_m (\mathbf{Y} \odot \mathbf{B}) \end{bmatrix} + \epsilon^3 h.o.t.\,,$$

Rearranging the terms, we have:

$$\frac{\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) + \mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B})) - 2\, \mathbf{f}(\mathbf{Y})}{\epsilon^2} = \begin{bmatrix} (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_1 (\mathbf{Y} \odot \mathbf{B}) \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_2 (\mathbf{Y} \odot \mathbf{B}) \\ \vdots \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_m (\mathbf{Y} \odot \mathbf{B}) \end{bmatrix} + \epsilon^3 h.o.t.\,,$$

Taking inner product with $(\mathbf{Y} \odot \mathbf{B})$ and computing expectation w.r.t B gives:

$$\begin{aligned} &\mathbb{E}_B \left\{ (\mathbf{Y} \odot \mathbf{B})^{\mathbf{T}} \left(\frac{\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) - \mathbf{2f}(\mathbf{Y}) + \mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B}))}{\epsilon^{\mathbf{2}}} \right) \right\} \\ &= \mathbb{E}_B \left\{ (\mathbf{Y} \odot \mathbf{B})^T \begin{bmatrix} (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_1 (\mathbf{Y} \odot \mathbf{B}) \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_2 (\mathbf{Y} \odot \mathbf{B}) \\ \vdots \\ (\mathbf{Y} \odot \mathbf{B})^T \mathbf{H}_m (\mathbf{Y} \odot \mathbf{B}) \end{bmatrix} + \epsilon^3 h.o.t. \right\}, \\ &= \mathbb{E}_B \left\{ \sum_i \sum_j \sum_k [\mathbf{H}_i]_{jk} Y_i Y_j Y_k b_i b_j b_k + \epsilon^3 h.o.t. \right\}. \end{aligned} \tag{14}$$

Since for $\mathbf{B}$, the entries are i.i.d with moments $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = \frac{-1}{5}$, Eq. 14 reduces to:

$$\mathbb{E}_B \left\{ (\mathbf{Y} \odot \mathbf{B})^{\mathbf{T}} \left(\frac{\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) - \mathbf{2f}(\mathbf{Y}) + \mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B}))}{\epsilon^{\mathbf{2}}} \right) \right\} = -\frac{1}{5} \sum_i [\mathbf{H}_i]_{ii} Y_i^3 + \epsilon^3 h.o.t.. \tag{15}$$

Taking the limit with $\epsilon \to 0$ vanishes the higher-order terms. With a minor re-arrangement, we can write Eq. 16 as:

$$\lim_{\epsilon \to 0} \mathbb{E}_B \left\{ -5(\mathbf{Y} \odot \mathbf{B})^{\mathbf{T}} \left(\frac{\mathbf{f}(\mathbf{Y} + \epsilon(\mathbf{Y} \odot \mathbf{B})) - \mathbf{2f}(\mathbf{Y}) + \mathbf{f}(\mathbf{Y} - \epsilon(\mathbf{Y} \odot \mathbf{B}))}{\epsilon^{\mathbf{2}}} \right) \right\} = \sum_i [\mathbf{H}_i]_{ii} Y_i^3 \tag{16}$$

Which completes the proof. Note that we write $[\mathbf{H}_i]_{ii}$ as $[\mathbf{H}_\mathbf{f}]_{iii}$ in the main paper to make explicit the fact that the Hessian is that of the function $\mathbf{f}$. □

Just like the $p = 1$ term, to compute this expectation, we utilize K realizations of $\mathbf{B}$ and n variations of ϵ. Hence, the estimate is evaluated as:

$$\sum_i [\mathbf{H}_\mathbf{f}]_{iii} Y_i^3 = \frac{1}{nK} \sum_{r=1}^{K} \sum_{s=1}^{n} \left\{ -5(\mathbf{Y} \odot \mathbf{B_r})^{\mathbf{T}} \left(\frac{\mathbf{f}(\mathbf{Y} + \epsilon_\mathbf{s}\,(\mathbf{Y} \odot \mathbf{B_r})) - \mathbf{2f}(\mathbf{Y}) + \mathbf{f}(\mathbf{Y} - \epsilon_\mathbf{s}\,(\mathbf{Y} \odot \mathbf{B_r}))}{\epsilon_\mathbf{s}^{\mathbf{2}}} \right) \right\}. \tag{17}$$

For all our experiments $n = K = 1$ yields reasonably good approximation with low computational complexity.

4. REFERENCES

[1] Chandra Sekhar Seelamantula and Thierry Blu, "Image denoising in multiplicative noise," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 1528–1532.