

# ON PROPERTIES OF THE POSTERIOR IN DENOISING: APPLICATION TO UNCERTAINTY QUANTIFICATION

**Hila Manor**

Faculty of Electrical and Computer Engineering  
Technion - Israel Institute of Technology  
[hila.manor@campus.technion.ac.il](mailto:hila.manor@campus.technion.ac.il)

**Tomer Michaeli**

Faculty of Electrical and Computer Engineering  
Technion – Israel Institute of Technology  
[tomer.m@ee.technion.ac.il](mailto:tomer.m@ee.technion.ac.il)

## ABSTRACT

Denoisers play a central role in many applications, from noise suppression in low-grade imaging sensors, to empowering score-based generative models. The latter category of methods makes use of Tweedie’s formula, which links the posterior mean in Gaussian denoising (*i.e.*, the minimum MSE denoiser) with the score of the data distribution. Here, we derive a fundamental relation between the higher-order central moments of the posterior distribution, and the higher-order derivatives of the posterior mean. We harness this result for uncertainty quantification of pre-trained denoisers. Particularly, we show how to efficiently compute the principal components of the posterior distribution for any desired region of an image, as well as to approximate the full marginal distribution along those (or any other) one-dimensional directions. Our method is fast and memory efficient, as it does not explicitly compute or store the high-order moment tensors and it requires no training or fine tuning of the denoiser. Code is available [here](#).

## 1 INTRODUCTION

Denoisers serve as key ingredients in solving a wide range of tasks. Indeed, along with their traditional use for noise suppression (Krull et al., 2019; Liang et al., 2021; Zhang et al., 2017a; 2021), the last decade has seen a steady increase in their use for solving other tasks. For example, the plug-and-play method (Venkatakrishnan et al., 2013) demonstrated how a denoiser can be used in an iterative manner to solve arbitrary inverse problems (*e.g.*, deblurring, inpainting). This approach was adopted and extended by many, and has led to state-of-the-art results on various restoration tasks (Brifman et al., 2016; Romano et al., 2017; Ticer & Giryes, 2018; Zhang et al., 2017b). Similarly, the denoising score-matching work (Vincent, 2011) showed how a denoiser can be used for constructing a generative model. This approach was later improved (Song & Ermon, 2019), and highly related ideas (originating from (Sohl-Dickstein et al., 2015)) served as the basis for diffusion models (Ho et al., 2020), which now achieve state-of-the-art results on image generation.

Many of the uses of denoisers rely on Tweedie’s formula (often attributed to Miyasawa et al. (1961), Efron (2011), and Stein (1981)) which connects between the MSE-optimal denoiser for white Gaussian noise, and the score function (the gradient of the log-probability) of the data distribution. The MSE-optimal denoiser corresponds to the posterior mean of the clean signal conditioned on the noisy signal. Therefore, Tweedie’s formula in fact links between the first posterior moment and the score of the data. A similar relation holds between the second posterior moment (*i.e.*, the posterior covariance) and the second-order score (*i.e.*, the Hessian of the log-probability) (Gribonval, 2011), which is in turn associated with the derivative (*i.e.*, Jacobian) of the posterior moment. Several recent works used this relation to quantify uncertainty in denoising (Meng et al., 2021), as well as to improve score-based generative models (Dockhorn et al., 2022; Lu et al., 2022; Meng et al., 2021; Mou et al., 2021; Sabanis & Zhang, 2019).

In this paper we derive a relation between higher-order posterior central moments and higher-order derivatives of the posterior mean in Gaussian denoising. Our result provides a simple mechanism that, given the MSE-optimal denoiser function and its derivatives at some input, allows determining the entire posterior distribution of clean signals for that particular noisy input (under mild condi-

tions). Additionally, we prove a similar result for the posterior distribution of the projection of the denoised output onto a one-dimensional direction.

We leverage our results for uncertainty quantification in Gaussian denoising by employing a pre-trained denoiser. Specifically, we show how our results allow computing the top eigenvectors of the posterior covariance (*i.e.*, the posterior principal components) for any desired region of the image. We further use our results for approximating the entire posterior distribution along each posterior principal direction. As we show, this allows extracting valuable information on the uncertainty in the restoration. Our approach requires access only to the forward pass of the pre-trained denoiser and is thus advantageous over previous uncertainty quantification methods. Particularly, it is training-free, fast, memory efficient, and applicable to high-resolution images. We illustrate our approach with several pre-trained denoisers on multiple domains.

## 2 RELATED WORK

Many works studied the theoretical properties of MSE-optimal denoisers for signals contaminated by additive white Gaussian noise. Perhaps the most well known result is Tweedie’s formula (Efron, 2011; Miyasawa et al., 1961; Stein, 1981), which connects between the MSE-optimal denoiser and the score function of noisy signals. Another interesting property, shown by Gribonval (2011), is that the MSE-optimal denoiser can be interpreted as a maximum-a-posteriori (MAP) estimator, but with a possibly different prior. The work most closely related to ours is that of Meng et al. (2021), who studied the estimation of high-order scores. Specifically, they derived a relation between the high-order posterior non-central moments in a Gaussian denoising task, and the high-order scores of the distribution of noisy signals. They discussed how these relations can be used for learning high-order scores of the data distribution. But due to the large memory cost of storing high-order moment tensors, and the associated computational cost during training and inference, they trained only second-order score models and only on small images (up to  $32 \times 32$ ). They used these models for predicting the posterior covariance in denoising tasks, as well as for improving the mixing speed of Langevin dynamics sampling. Their result is based on a recursive relation, which they derived, between the high-order derivatives of the posterior mean and the high-order *non-central* moments of the posterior distribution in Gaussian denoising. Specifically, they showed that the non-central posterior moments  $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \dots$ , admit a recursion of the form  $\mathbf{m}_{k+1} = f(\mathbf{m}_k, \nabla \mathbf{m}_k, \mathbf{m}_1)$ .

In many settings, *central* moments are rather preferred over their non-central counterparts. Indeed, they are more numerically stable and relate more intuitively to uncertainty quantification (being directly linked to variance, skewness, kurtosis, etc.). Unfortunately, the result of (Meng et al., 2021) does not trivially translate into a useful relation for central moments. Specifically, one could use the fact that the  $k$ th central moment,  $\mu_k$ , can be expressed in terms of  $\{\mathbf{m}_j\}_{j=1}^k$ , and that each  $\mathbf{m}_j$  can be written in terms of  $\{\mu_i\}_{i=1}^j$ . But naively substituting these relations into the recursion of Meng et al. (2021) leads to an expression for  $\mu_k$  that includes all lower-order central-moments and their high-order derivatives. Here, we manage to prove a very simple recursive form for the central moments, which takes the form  $\mu_{k+1} = f(\mu_k, \nabla \mu_k, \mu_2)$ . Another key contribution, which we present beyond the framework studied by Meng et al. (2021), relates to marginal posterior distributions along arbitrary cross-sections. Specifically, we prove that the central posterior moments of any low-dimensional projection of the signal, also satisfy a similar recursion. Importantly, we show how these relations can serve as very powerful tools for uncertainty quantification in denoising tasks.

Uncertainty quantification has drawn significant attention in the context of image restoration. Many works focused on per-pixel uncertainty prediction (Angelopoulos et al., 2022; Gal & Ghahramani, 2016; Horwitz & Hoshen, 2022; Meng et al., 2021; Oala et al., 2020), which neglects correlations between the uncertainties of different pixels in the restored image. Recently, several works forayed into more meaningful notions of uncertainty, which allow to reason about semantic variations (Kutiel et al., 2023; Sankaranarayanan et al., 2022). However, existing methods either require a pre-trained generative model with a disentangled latent space (*e.g.*, StyleGAN (Karras et al., 2020) for face images) or, like many of their per-pixel counterparts, require training. Here we present a training-free, computationally efficient, method that only requires access to a pre-trained denoiser.

### 3 MAIN THEORETICAL RESULT

We now present our main theoretical result, starting with scalar denoising and then extending the discussion to the multivariate setting. The scalar case serves two purposes. First, it provides intuition. But more importantly, the formulae for moments of orders higher than three are different for the univariate and multivariate settings, and therefore the two cases require separate treatment.

#### 3.1 THE UNIVARIATE CASE

Consider the univariate denoising problem

$$y = x + n, \quad (1)$$

where  $x$  is a scalar random variable with probability density function  $p_x$  and the noise  $n \sim \mathcal{N}(0, \sigma^2)$  is statistically independent of  $x$ . The goal in denoising is to provide a prediction  $\hat{x}$  of  $x$ , which is a function of the measurement  $y$ . It is well known that the predictor minimizing the MSE,  $\mathbb{E}[(x - \hat{x})^2]$ , is the posterior mean of  $x$  given  $y$ . Specifically, given a particular measurement  $y = y$ , the MSE-optimal estimate is the first moment of the posterior density  $p_{x|y}(\cdot|y)$ , which we denote by

$$\mu_1(y) = \mathbb{E}[x|y = y]. \quad (2)$$

While optimal in the MSE sense, the posterior mean provides very partial knowledge on the possible values that  $x$  could take given that  $y = y$ . More information is encoded in higher-order moments of the posterior. For example, the posterior variance provides a measure of uncertainty about the MSE-optimal prediction, the posterior third moment provides knowledge about the skewness of the posterior distribution, and the posterior fourth moment can already reveal a bimodal behavior.

Let us denote the higher-order posterior central moments by

$$\mu_k(y) = \mathbb{E}[(x - \mu_1(y))^k | y = y], \quad k \geq 2. \quad (3)$$

Our key result is that knowing the posterior mean function  $\mu_1(\cdot)$  and its derivatives at  $y$  can be used to recursively compute all higher-order posterior central moments at  $y$  (see proof in App. A).

**Theorem 1** (Posterior moments in univariate denoising). *In the scalar denoising setting of (1), the high-order posterior central moments of  $x$  given  $y$  satisfy the recursion*

$$\begin{aligned} \mu_2(y) &= \sigma^2 \mu'_1(y), \\ \mu_3(y) &= \sigma^2 \mu''_1(y), \\ \mu_{k+1}(y) &= \sigma^2 \mu'_k(y) + k\mu_{k-1}(y)\mu_2(y), \quad k \geq 3. \end{aligned} \quad (4)$$

Thus,  $\mu_{k+1}(y)$  is uniquely determined by  $\mu_1(y), \mu'_1(y), \mu''_1(y), \dots, \mu_1^{(k)}(y)$ .

Figure 1 illustrates this result via a simple example. Here, the distribution of  $x$  is a mixture of two Gaussians. The left pane depicts the posterior density  $p_{x|y}(\cdot|y)$  as well as the posterior mean function  $\mu_1(\cdot)$ . We focus on the measurement  $y = y^*$ , shown as a vertical dashed line, for which the posterior  $p_{x|y}(\cdot|y^*)$  is bimodal (right pane). This property cannot be deduced by merely examining the MSE-optimal estimate  $\mu_1(y^*)$ . However, this information does exist in the derivatives of  $\mu_1(\cdot)$  at  $y^*$ . To demonstrate this, we numerically differentiated  $\mu_1(\cdot)$  at  $y^*$ , used the first three derivatives to extract the first four posterior moments using Theorem 1, and computed the maximum entropy distribution that matches those moments (Botev & Kroese, 2011). As can be seen, this already provides a good approximation of the general shape of the posterior (dashed red line).

Theorem 1 has several immediate implications. First, it is well known that if the moments do not grow too fast, then they uniquely determine the underlying distribution (Lin, 2017). This is the case e.g., for distributions with a compact support and is thus relevant to images, whose pixel values typically lie in  $[0, 1]$ . For such settings, Theorem 1 implies that knowing the posterior mean at the neighborhood of some point  $y$ , allows determining the entire posterior distribution for that point. A second interesting observation, is that Theorem 1 can be evoked to show that the posterior is Gaussian whenever all high-order derivatives of  $\mu_1(\cdot)$  vanish (see proof in App. F).

**Corollary 1.** *Assume that  $\mu_1^{(k)}(y^*) = 0$  for all  $k > 1$ . Then the posterior  $p_{x|y}(\cdot|y^*)$  is Gaussian.*

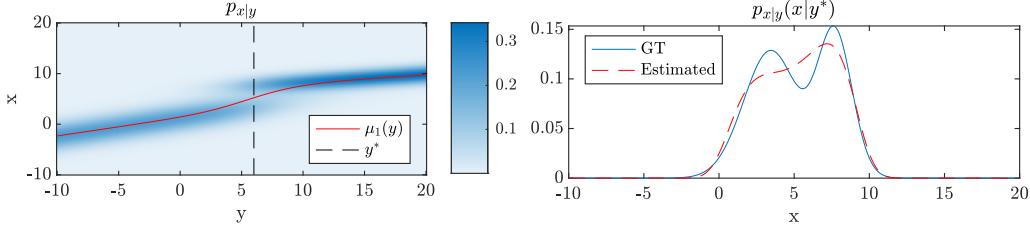


Figure 1: **Recovering posteriors in univariate denoising.** The left pane shows the posterior distribution  $p_{x|y}(\cdot|\cdot)$  and the posterior mean function  $\mu_1(\cdot)$  for the scalar Gaussian denoising task (1). On the right we plot the posterior distribution of  $x$  given that  $y = y^*$ , along with an estimate of that distribution, which we obtain by analyzing the denoiser function  $\mu_1(\cdot)$  at the vicinity of  $y^*$ . Specifically, this estimate corresponds to the maximum entropy distribution that matches the first four moments, which are obtained from Theorem 1 by numerically approximating  $\mu'_1(y^*)$ ,  $\mu''_1(y^*)$ ,  $\mu'''_1(y^*)$ .

### 3.2 THE MULTIVARIATE CASE

We now move on to treat the multivariate denoising problem. Here  $\mathbf{x}$  is a random vector taking values in  $\mathbb{R}^d$ , the noise  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$  is a white multivariate Gaussian vector that is statistically independent of  $\mathbf{x}$ , and the noisy observation is

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \quad (5)$$

As in the scalar setting, given a noisy measurement  $\mathbf{y} = \mathbf{y}$ , we are interested in the posterior distribution  $p_{\mathbf{x}|\mathbf{y}}(\cdot|\mathbf{y})$ . The MSE-optimal denoiser is, again, the the first-order moment of this distribution,

$$\boldsymbol{\mu}_1(\mathbf{y}) = \mathbb{E}[\mathbf{x} | \mathbf{y} = \mathbf{y}], \quad (6)$$

which is a  $d$  dimensional vector. The second-order central moment is the posterior covariance

$$\boldsymbol{\mu}_2(\mathbf{y}) = \text{Cov}(\mathbf{x} | \mathbf{y} = \mathbf{y}), \quad (7)$$

which is a  $d \times d$  matrix whose  $(i_1, i_2)$  entry is given by

$$[\boldsymbol{\mu}_2(\mathbf{y})]_{i_1, i_2} = \mathbb{E}[(\mathbf{x}_{i_1} - [\boldsymbol{\mu}_1(\mathbf{y})]_{i_1})(\mathbf{x}_{i_2} - [\boldsymbol{\mu}_1(\mathbf{y})]_{i_2}) | \mathbf{y} = \mathbf{y}]. \quad (8)$$

For any  $k \geq 3$ , the posterior  $k$ th-order central moment is a  $d \times \dots \times d$  array with  $k$  indices (a  $k$ th order tensor), whose component at multi-index  $(i_1, \dots, i_k)$  is given by

$$[\boldsymbol{\mu}_k(\mathbf{y})]_{i_1, \dots, i_k} = \mathbb{E}[(\mathbf{x}_{i_1} - [\boldsymbol{\mu}_1(\mathbf{y})]_{i_1}) \cdots (\mathbf{x}_{i_k} - [\boldsymbol{\mu}_1(\mathbf{y})]_{i_k}) | \mathbf{y} = \mathbf{y}]. \quad (9)$$

As we now show, similarly to the scalar case, having access to the MSE-optimal denoiser and its derivatives, allows to recursively compute all higher order posterior moments (see proof in App. B).

**Theorem 2** (Posterior moments in multivariate denoising). *Consider the multivariate denoising setting of (5) with dimension  $d \geq 2$ . For any  $k \geq 1$  and any  $k+1$  indices  $i_1, \dots, i_{k+1} \in \{1, \dots, d\}$ , the high-order posterior central moments of  $\mathbf{x}$  given  $\mathbf{y}$  satisfy the recursion*

$$\begin{aligned} [\boldsymbol{\mu}_2(\mathbf{y})]_{i_1, i_2} &= \sigma^2 \frac{\partial [\boldsymbol{\mu}_1(\mathbf{y})]_{i_1}}{\partial \mathbf{y}_{i_2}}, \\ [\boldsymbol{\mu}_3(\mathbf{y})]_{i_1, i_2, i_3} &= \sigma^2 \frac{\partial [\boldsymbol{\mu}_2(\mathbf{y})]_{i_1, i_2}}{\partial \mathbf{y}_{i_3}}, \\ [\boldsymbol{\mu}_{k+1}(\mathbf{y})]_{i_1, \dots, i_{k+1}} &= \sigma^2 \frac{\partial [\boldsymbol{\mu}_k(\mathbf{y})]_{i_1, \dots, i_k}}{\partial \mathbf{y}_{i_{k+1}}} + \sum_{j=1}^k [\boldsymbol{\mu}_{k-1}(\mathbf{y})]_{\ell_j} [\boldsymbol{\mu}_2(\mathbf{y})]_{i_j, i_{k+1}}, \quad k \geq 3, \end{aligned} \quad (10)$$

where  $\ell_j \triangleq (i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_k)$ . Thus,  $\boldsymbol{\mu}_{k+1}(\mathbf{y})$  is uniquely determined by  $\boldsymbol{\mu}_1(\mathbf{y})$  and by the derivatives up to order  $k$  of its elements with respect to the elements of the vector  $\mathbf{y}$ .

Note that the first line in (10) can be compactly written as

$$\boldsymbol{\mu}_2(\mathbf{y}) = \sigma^2 \frac{\partial \boldsymbol{\mu}_1(\mathbf{y})}{\partial \mathbf{y}}, \quad (11)$$

**Algorithm 1** Efficient computation of posterior principal components

---

**Input:**  $N$  (number of PCs),  $K$  (number of iterations),  $\mu_1(\cdot)$  (MSE-optimal denoiser),  $\mathbf{y}$  (noisy input),  $\sigma^2$  (noise variance),  $c \ll 1$  (linear approx. constant)

- 1: Initialize  $\{\mathbf{v}_0^{(i)}\}_{i=1}^N \leftarrow \mathcal{N}(0, \sigma^2 \mathbf{I})$
- 2: **for**  $k \leftarrow 1$  to  $K$  **do**
- 3:     **for**  $i \leftarrow 1$  to  $N$  **do**
- 4:          $\mathbf{v}_k^{(i)} \leftarrow \frac{1}{c} (\mu_1(\mathbf{y} + c\mathbf{v}_{k-1}^{(i)}) - \mu_1(\mathbf{y}))$
- 5:     **Q, R**  $\leftarrow$  QR\_DECOMPOSITION( $[\mathbf{v}_k^{(1)} \cdots \mathbf{v}_k^{(N)}]$ )
- 6:      $[\mathbf{v}_k^{(1)} \cdots \mathbf{v}_k^{(N)}] \leftarrow \mathbf{Q}$
- 7:      $\mathbf{v}^{(i)} \leftarrow \mathbf{v}_K^{(i)}$
- 8:      $\lambda^{(i)} \leftarrow \frac{\sigma^2}{c} \|\mu_1(\mathbf{y} + c\mathbf{v}_{K-1}^{(i)}) - \mu_1(\mathbf{y})\|$

---

where  $\frac{\partial \mu_1(\mathbf{y})}{\partial \mathbf{y}}$  denotes the Jacobian of  $\mu_1$  at  $\mathbf{y}$ . This suggests that, in principle, the posterior covariance of an MSE-optimal denoiser could be extracted by computing the Jacobian of the model using *e.g.*, automatic differentiation. However, in settings involving high-resolution images, even storing this Jacobian is impractical. In Sec. 4.1, we show how the top eigenvectors of  $\mu_2(\mathbf{y})$  (*i.e.*, the posterior principal components) can be computed without having to ever store  $\mu_2(\mathbf{y})$  in memory.

Moments of order greater than two pose an even bigger challenge, as they correspond to higher-order tensors. In fact, even if they could somehow be computed, it is not clear how they would be visualized in order to communicate the uncertainty of the prediction to a user. A practical solution could be to visualize the posterior distribution of the projection of  $\mathbf{x}$  onto some meaningful one-dimensional space. For example, one might be interested in the posterior distribution of  $\mathbf{x}$  projected onto one of the principal components of the posterior covariance. The question, however, is how to obtain the posterior moments of the projection of  $\mathbf{x}$  onto a deterministic  $d$ -dimensional vector  $\mathbf{v}$ .

Let us denote the first posterior moment of  $\mathbf{v}^\top \mathbf{x}$  (*i.e.*, its posterior mean) by  $\mu_1^\mathbf{v}(\mathbf{y})$ . This moment is given by the projection of the denoiser's output onto  $\mathbf{v}$ ,

$$\mu_1^\mathbf{v}(\mathbf{y}) = \mathbb{E} [\mathbf{v}^\top \mathbf{x} | \mathbf{y} = \mathbf{y}] = \mathbf{v}^\top \mathbb{E} [\mathbf{x} | \mathbf{y} = \mathbf{y}] = \mathbf{v}^\top \mu_1(\mathbf{y}). \quad (12)$$

Similarly, let us denote the  $k$ th order posterior central moment of  $\mathbf{v}^\top \mathbf{x}$  by

$$\mu_k^\mathbf{v}(\mathbf{y}) = \mathbb{E} [(\mathbf{v}^\top \mathbf{x} - \mu_1^\mathbf{v}(\mathbf{y}))^k | \mathbf{y} = \mathbf{y}], \quad k \geq 2. \quad (13)$$

As we show next, the scalar-valued functions  $\{\mu_k^\mathbf{v}(\mathbf{y})\}_{k=1}^\infty$  satisfy a recursion similar to (4) (see proof in App. C). In Sec. 5, we use this result for uncertainty visualization.

**Theorem 3** (Directional posterior moments in multivariate denoising). *Let  $\mathbf{v}$  be a deterministic  $d$ -dimensional vector. Then the posterior central moments of  $\mathbf{v}^\top \mathbf{x}$  are given by the recursion*

$$\begin{aligned} \mu_2^\mathbf{v}(\mathbf{y}) &= \sigma^2 D_{\mathbf{v}} \mu_1^\mathbf{v}(\mathbf{y}), \\ \mu_3^\mathbf{v}(\mathbf{y}) &= \sigma^2 D_{\mathbf{v}} \mu_2^\mathbf{v}(\mathbf{y}), \\ \mu_{k+1}^\mathbf{v}(\mathbf{y}) &= \sigma^2 D_{\mathbf{v}} \mu_k^\mathbf{v}(\mathbf{y}) + k \mu_{k-1}^\mathbf{v}(\mathbf{y}) \mu_2^\mathbf{v}(\mathbf{y}), \quad k \geq 3. \end{aligned} \quad (14)$$

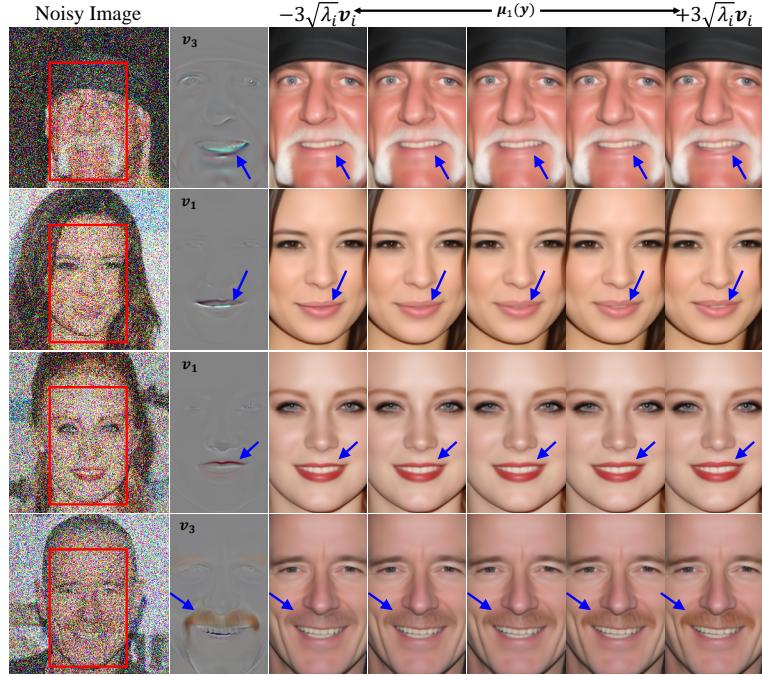
Here  $D_{\mathbf{v}} f(\mathbf{y})$  denotes the directional derivative of a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  in direction  $\mathbf{v}$  at  $\mathbf{y}$ .

## 4 APPLICATION TO UNCERTAINTY VISUALIZATION

We now discuss the applicability of our results in the context of uncertainty visualization. We start with efficient computation of posterior principal components, and then illustrate the approximation of marginal densities along those directions.

### 4.1 EFFICIENT COMPUTATION OF POSTERIOR PRINCIPAL COMPONENTS

The top eigenvectors of the posterior covariance,  $\mu_2(\mathbf{y})$ , capture the main modes of variation about the MSE-optimal prediction. Thus, as we illustrate below, they reveal meaningful information regarding the uncertainty of the restoration. Had we had access to the matrix  $\mu_2(\mathbf{y})$ , computing



**Figure 2: Computing posterior principal components for a pre-trained face denoising model.** For each noisy image  $\mathbf{y}$ , we depict one of the posterior PCs obtained with Alg. 1. To the right of that PC, we show the denoiser’s output,  $\mu_1(\mathbf{y})$ , and its perturbation along that PC. As can be seen, this visualization captures the denoiser’s uncertainty along semantically meaningful directions, such as the color of the moustache, the thickness of the lips, and the extent to which the mouth is open.

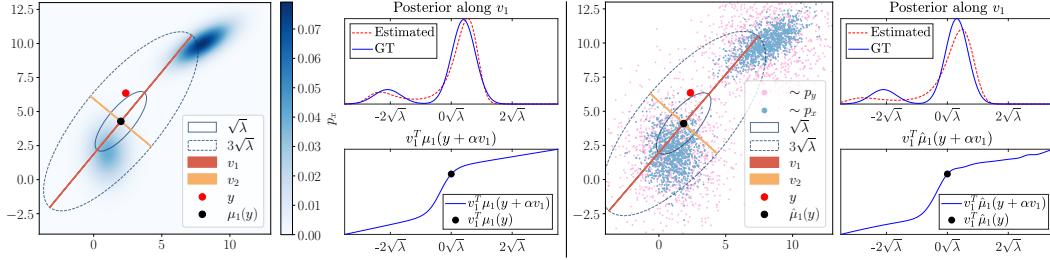
these top eigenvectors could be done using the subspace iteration method (Arbenz, 2016; Saad, 2011). This technique maintains a set of  $N$  vectors, which are repeatedly multiplied by  $\mu_2(\mathbf{y})$  and orthonormalized using the QR decomposition. Unfortunately, storing the full covariance matrix is commonly impractical. To circumvent the need for doing so, we recall from (11) that  $\mu_2(\mathbf{y})$  corresponds to the Jacobian of the denoiser  $\mu_1(\mathbf{y})$ . Thus, every iteration of the subspace method corresponds to a Jacobian-vector dot-product. For neural denoisers, such products can be calculated using automatic differentiation (Dockhorn et al., 2022). However, this requires computing a backward pass through the model in each iteration, which can become computationally demanding for large images<sup>1</sup>. Instead, we propose to use the linear approximation

$$\frac{\partial \mu_1(\mathbf{y})}{\partial \mathbf{y}} \mathbf{v} \approx \frac{\mu_1(\mathbf{y} + c\mathbf{v}) - \mu_1(\mathbf{y})}{c}, \quad (15)$$

which holds for any  $\mathbf{v} \in \mathbb{R}^d$  when  $c \in \mathbb{R}$  is sufficiently small. This allows applying the subspace iteration using only forward passes through the denoiser, as summarized in Alg. 1. As we show in App. H, this approximation has a negligible effect on the calculated eigenvectors, but leads *e.g.*, to a  $6\times$  reduction in memory footprint for a  $80 \times 92$  patch with the SwinIR denoiser (Liang et al., 2021).

Figure 2 illustrates this technique in the context of denoising of face images contaminated by white Gaussian noise with standard deviation  $\sigma = 122$ . We use the denoiser from (Baranchuk et al., 2022), which was trained as part of a DDPM model (Ho et al., 2020) on the FFHQ dataset (Karras et al., 2019). Note that here we use it as a plain denoiser (as used within a single timestep of the DDPM). We showcase examples from the CelebAMask-HQ dataset (Lee et al., 2020). As can be seen, different posterior principal components typically capture uncertainty in different localized regions of the image. Note that this approach can be applied to any region-of-interest within the image, chosen by the user at test time. This is in contrast to a model that is trained to predict a low-rank approximation of the covariance, as in (Meng et al., 2021). Such a model is inherently limited

<sup>1</sup>Note that backward passes for whole images are also often avoided during training of neural denoisers. Indeed, typical training procedures use limited-sized crops.



**Figure 3: Computing marginals along principal components.** On the left, we show the prior  $p_{\mathbf{x}}$  as a heatmap, a noisy sample  $\mathbf{y}$  (red), the corresponding MSE-optimal estimate  $\mu_1(\mathbf{y})$  (black), and the two principal axes, computed using Alg. 1. Here, we used the closed form for  $\mu_1(\mathbf{y})$ . The second pane shows the marginal posterior distribution along the first principal component, computed both using our proposed procedure (dashed red), and by using the closed-form solution (solid blue). On the right we show the same experiment, but with a simple neural network trained on data samples.

to the specific input size on which it was trained, and cannot be manipulated at test time to produce eigenvectors corresponding to some user-chosen region (cropping a patch from an eigenvector is not equivalent to computing the eigenvector of the corresponding patch in the image).

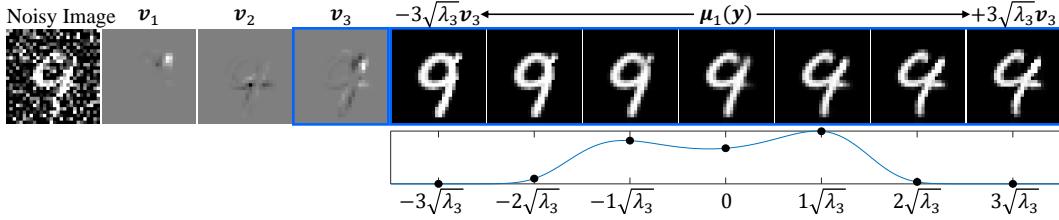
#### 4.2 ESTIMATION OF MARGINAL DISTRIBUTIONS ALONG CHOSEN DIRECTIONS

A more fine-grained characterization of the posterior can be achieved by using higher-order moments along the principal directions. These can be calculated using Theorem 3, through (high-order) numerical differentiation of the one-dimensional function  $f(\alpha) = \mathbf{v}^\top \mu_1(\mathbf{y} + \alpha \mathbf{v})$  at  $\alpha = 0$ . Once we obtain all moments up to some order, we compute the probability distribution with maximum entropy that fits those moments. In practice, we compute derivatives up to third order, which allows us to obtain all moments up to order four.

Figure 3 illustrates this approach on a two-dimensional Gaussian mixture example with a noise level of  $\sigma = 2$ . On the left, we show a heatmap corresponding to  $p_{\mathbf{x}}(\cdot)$ , as well as a noisy input  $\mathbf{y}$  (red point) and its corresponding MSE-optimal estimate (black point). The two axes of the ellipse are the posterior principal components computed using Alg. 1 using numerical differentiation of the closed-form expression of the denoiser (see App. E). The bottom plot on the second pane shows the function  $f(\alpha)$  corresponding to the largest eigenvector. We numerically computed its derivatives up to order three at  $\alpha = 0$  (black point), from which we estimated the moments up to order four according to Theorem 3. The top plot on that pane shows the ground-truth posterior distribution of  $v_1^\top \mathbf{x}$ , along with the maximum entropy distribution computed from the moments. The right half of the figure shows the same experiment only with a neural network that was trained on pairs of noisy (pink) samples and their clean (blue) counterparts. This denoiser comprises 5 layers with (100, 200, 200, 100) hidden features and SiLU (Hendrycks & Gimpel, 2016) activation units. We trained the network using Adam (Kingma & Ba, 2015) for 300 epochs, with a learning rate of 0.005.

Figure 4 illustrates the approach on a handwritten digit from the MNIST (LeCun, 1998) dataset. Here, we train and use a simple CNN with 10 layers of 64 channels, separated by ReLU activation layers followed by batch normalization layers. As can be seen, fitting the maximum entropy distribution reveals more than just the main modes of variation, as it also reveals the likelihood of each reconstruction along that direction. It is instructive to note that although the two extreme reconstructions,  $\mu_1(\mathbf{y}) \pm \sqrt{\lambda_3} \mathbf{v}_3$ , look realistic, they are not probable given the noisy observation. This is the reason their corresponding estimated posterior density is nearly zero.

Our theoretical analysis applies to non-blind denoising, in which  $\sigma$  is known. However, we empirically show in Sec. 5 and Fig. 5 that using an estimated  $\sigma$  is also sufficient for obtaining qualitatively plausible results. This can either be obtained from a noise estimation method (Chen et al., 2015) or even from the naive estimate  $\hat{\sigma}^2 = \frac{1}{d} \|\mu_1(\mathbf{y}) - \mathbf{y}\|^2$ , where  $\mu_1(\mathbf{y})$  is the output of a blind denoiser. Here we use the latter. We further discuss the impact of using an estimated  $\sigma$  in App. I.



**Figure 4: Uncertainty quantification for denoising a handwritten digit.** The first three PCs corresponding to the noisy image are shown on the left. On the right, images along the third PC, marked in blue, are shown, together with the marginal posterior distribution we estimated for that direction. The two modes of the possible restoration, corresponding to the digits 4 and 9, can clearly be seen as peaks in the marginal posterior distribution, whereas the MSE-optimal restoration in the middle is obviously less likely.

## 5 EXPERIMENTS

We conduct experiments with our proposed approach for uncertainty visualization and marginal posterior distribution estimation on additional real data in multiple domains using different models.

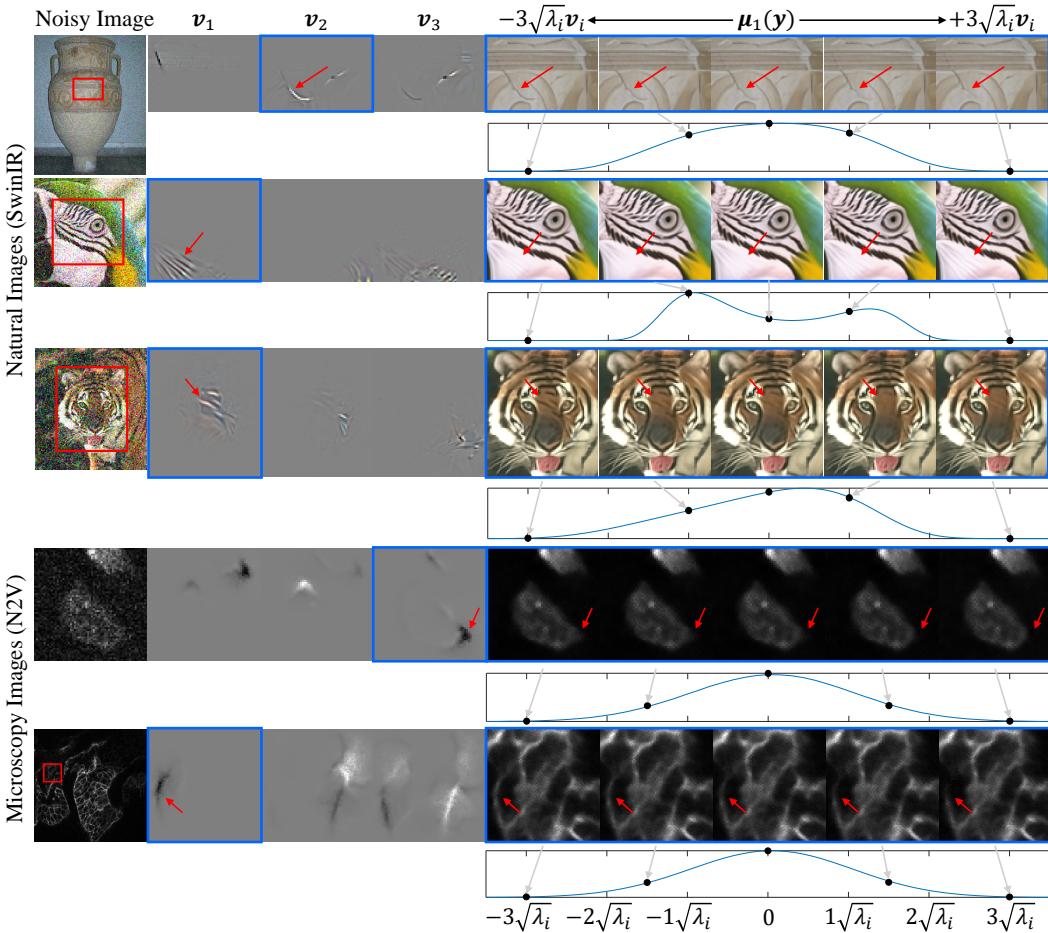
We showcase our method on the MNIST dataset, natural images, human faces, and on images from the microscopy domain. For natural images, we use SwinIR (Liang et al., 2021) that was pre-trained on 800 DIV2K (Agustsson & Timofte, 2017) images, 2650 Flickr2k (Lim et al., 2017) images, 400 BSD500 (Arbelaez et al., 2010) images and 4,744 WED (Ma et al., 2016) images, with patch sizes  $128 \times 128$  and window size  $8 \times 8$ . We experiment with two SwinIR models, trained separately for noise levels  $\sigma = \{25, 50\}$ , and showcase examples on test images from the CBSD68 (Martin et al., 2001) and Kodak (Franzen, 1999) datasets. For the medical and microscopy domain we use Noise2Void (Krull et al., 2019), trained and tested for blind-denoising on the FMD dataset (Zhang et al., 2019) in the unsupervised manner described by Krull et al. (2020). The FMD dataset was collected using real microscopy imaging, and as such its noise is most probably not precisely white nor Gaussian, and the noise level is unknown in essence (the ground truth images are considered as the average of 50 burst images). Accordingly, N2V is a blind-denoiser, and we have no access to the “real”  $\sigma$ , therefore, for this dataset we used an estimated  $\sigma$  in our method, as described in Sec. 4.2.

Examples for the different domains can be seen in Figs. 2, 4, and 5. As can be seen, in all cases, our approach captures interesting uncertainty directions. For natural images, those include cracks, wrinkles, eye colors, stripe shapes, etc. In the biological domain, visualizations reveal uncertainty in the size and morphology of cells, as well as in the (in)existence of septum. Those constitute important geometric features in cellular analysis. More examples can be found in App. J.

One limitation of the proposed method is that it relies on high-order numerical differentiation. As this approximation can be unstable with low-precision computation, we use double precision during the forward pass of the networks. Another method that can be used to mitigate this is to fit a low degree polynomial to  $f(\alpha) = \mathbf{v}^\top \boldsymbol{\mu}_1(\mathbf{y} + \alpha \mathbf{v})$  around the point of derivation,  $\alpha = 0$ , and then use the smooth polynomial fit for the high-order derivatives calculation. Empirically we found the polynomial fitting to also be sensitive, highly-dependant on the choice of the polynomial degree and the fitted range. This caused bad fits even for the simple two-component GMM example, whereas the numerical derivatives approximations worked better.

## 6 CONCLUSION

Denoisers constitute fundamental ingredients in a variety of problems. In this paper we derived a relation in the denoising problem between higher-order derivatives of the posterior mean to higher-order posterior central moments. These results were then used in the application of uncertainty visualisation of pre-trained denoisers. Specifically, we proposed a method for efficiently computing the principal components of the posterior distribution, in any chosen region of an image. Additionally, we presented a scheme to use higher-order moments to estimate the full marginal distribution along



**Figure 5: Uncertainty quantification for natural image denoising using SwinIR (top) and microscopy image denoising using N2V (bottom).** In each row, the first three PCs corresponding to the noisy image are shown on the left, and one is marked in blue. On the right, images along the marked PC are shown above the marginal posterior distribution estimated for this direction. The PCs show the uncertainty along meaningful directions, such as the existence of cracks on an old vase and changes in the tiger’s stripes, as well as the sizes of cells and the existence of septum, which constitute important geometric features in cellular analysis.

any one-dimensional direction. Finally, we demonstrated our method on multiple denoisers across different domains. Our method allows examining semantic directions of uncertainty by using only pre-trained denoisers, in a fast and memory efficient way. While the theoretical basis of our method holds only for denoising additive white Gaussian noise, we show empirically that it provides qualitatively satisfactory results also in blind Poisson-Gaussian denoising. As deep learning becomes more intertwined in our daily lives, uncertainty quantification becomes an important tool for trustworthy interpretable models. Our method can be used to help visualize the full spectrum of possibilities in denoised restorations, and help take credible decisions, *e.g.*, in healthcare applications.

## REPRODUCIBILITY STATEMENT

As part of the ongoing effort to make the field of deep learning more reproducible and open, we publish our code at <https://hilamanor.github.io/GaussianDenoisingPosterior/>. The repository includes scripts to regenerate all figures. Researchers that want to re-implement the code from scratch can use Alg. 1 and our published code as guidelines. In addition, we provide full

and detailed proofs for all claims in the paper in Appendices A, B, C, E, and F of the supplementary material. Finally, we provide in Appendix D a translation from our notation to the notation of Meng et al. (2021) to allow future researchers to use both methods conveniently.

## REFERENCES

- Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017. URL <https://data.vision.ee.ethz.ch/cvl/DIV2K/>.
- Anastasios N Angelopoulos, Amit Pal Kohli, Stephen Bates, Michael Jordan, Jitendra Malik, Thayer Alshaabi, Srigokul Upadhyayula, and Yaniv Romano. Image-to-image regression with distribution-free uncertainty quantification and applications in imaging. In *International Conference on Machine Learning*, pp. 717–730. PMLR, 2022.
- Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010. URL <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>.
- Peter Arbenz. Lecture notes on solving large scale eigenvalue problems. 2016.
- Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Label-efficient semantic segmentation with diffusion models. In *International Conference on Learning Representations*, 2022.
- Zdravko I Botev and Dirk P Kroese. The generalized cross entropy method, with applications to probability density estimation. *Methodology and Computing in Applied Probability*, 13(1):1–27, 2011.
- Alon Brifman, Yaniv Romano, and Michael Elad. Turning a denoiser into a super-resolver using plug and play priors. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1404–1408. IEEE, 2016.
- Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 477–485, 2015.
- Tim Dockhorn, Arash Vahdat, and Karsten Kreis. GENIE: Higher-order denoising diffusion solvers. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022.
- Bradley Efron. Tweedie’s formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.
- Rich Franzen. Kodak lossless true color image suite. *source: http://r0k.us/graphics/kodak*, 1999.
- Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.
- Rémi Gribonval. Should penalized least squares regression be interpreted as maximum a posteriori estimation? *IEEE Transactions on Signal Processing*, 59(5):2405–2410, 2011.
- Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- Eliahu Horwitz and Yedid Hoshen. Conffusion: Confidence intervals for diffusion models. *arXiv preprint arXiv:2211.09795*, 2022.

- Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019. URL <https://github.com/NVlabs/fhq-dataset>.
- Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33:12104–12114, 2020.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 3, 2015.
- Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2129–2137, 2019.
- Alexander Krull, Tomáš Vičar, Mangal Prakash, Manan Lalit, and Florian Jug. Probabilistic noise2void: Unsupervised content-aware denoising. *Frontiers in Computer Science*, 2:5, 2020.
- Gilad Kutiel, Regev Cohen, Michael Elad, Daniel Freedman, and Ehud Rivlin. Conformal prediction masks: Visualizing uncertainty in medical imaging. In *ICLR 2023 Workshop on Trustworthy Machine Learning for Healthcare*, 2023.
- Yan LeCun. The MNIST database of handwritten digits. 1998. URL <http://yann.lecun.com/exdb/mnist/>.
- Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5549–5558, 2020.
- Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.
- Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. URL <https://github.com/limbee/NTIRE2017>.
- Gwo Dong Lin. Recent developments on the moment problem. *Journal of Statistical Distributions and Applications*, 4(1):5, 2017.
- Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Maximum likelihood training for score-based diffusion odes by high order denoising score matching. In *International Conference on Machine Learning*, pp. 14429–14460. PMLR, 2022.
- Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016. URL <https://ece.uwaterloo.ca/~k29ma/exploration/>.
- David Martin, Charless Fowlkes, Tal D., and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pp. 416–423. IEEE, 2001. URL <https://github.com/clausmichele/CBSD68-dataset>.
- Chenlin Meng, Yang Song, Wenzhe Li, and Stefano Ermon. Estimating high order gradients of the data distribution by denoising. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 25359–25369. Curran Associates, Inc., 2021.
- Koichi Miyasawa et al. An empirical bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist.*, 38(181-188):1–2, 1961.

- Wenlong Mou, Yi-An Ma, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. High-order langevin diffusion yields an accelerated MCMC algorithm. *Journal of Machine Learning Research*, 22(42):1–41, 2021.
- Luis Oala, Cosmas Heiß, Jan Macdonald, Maximilian März, Wojciech Samek, and Gitta Kutyniok. Interval neural networks: Uncertainty scores. *arXiv preprint arXiv:2003.11566*, 2020.
- Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.
- Yousef Saad. *Numerical methods for large eigenvalue problems: revised edition*. SIAM, 2011.
- Sotirios Sabanis and Ying Zhang. Higher order langevin monte carlo algorithm. *Electronic Journal of Statistics*, 13(2):3805–3850, 2019.
- Swami Sankaranarayanan, Anastasios Nikolas Angelopoulos, Stephen Bates, Yaniv Romano, and Phillip Isola. Semantic uncertainty intervals for disentangled latent spaces. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022.
- Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.
- Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.
- Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.
- Tom Tirer and Raja Giryes. Image restoration by iterative denoising and backward projections. *IEEE Transactions on Image Processing*, 28(3):1220–1234, 2018.
- Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948. IEEE, 2013.
- Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.
- Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017a.
- Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938, 2017b.
- Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.
- Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11710–11718, 2019. URL <https://github.com/yinhaoz/denoising-fluorescence>.