Anisotropic Nonlocal Means Denoising

Arian Maleki[1,*], Manjari Narayan, Richard G. Baraniuk[**]

Department of Electrical and Computer Engineering
Rice University

Abstract

It has recently been proved that the popular nonlocal means (NLM) denoising algorithm does not optimally denoise images with sharp edges. Its weakness lies in the isotropic nature of the neighborhoods it uses to set its smoothing weights. In response, in this paper we introduce several theoretical and practical *anisotropic nonlocal means* (ANLM) algorithms and prove that they are near minimax optimal for edge-dominated images from the Horizon class. On real-world test images, an ANLM algorithm that adapts to the underlying image gradients outperforms NLM by a significant margin.

Keywords: Denoising, nonlocal means, minimax risk, anisotropy

1. Introduction

Image denoising is a fundamental primitive in image processing and computer vision. Denoising algorithms have evolved from the classical linear and median filters to more modern schemes like total variation denoising [1], wavelet thresholding [2], and bilateral filters [3–6].

A particularly successful denoising scheme is the *nonlocal means* (NLM) algorithm [7], which estimates each pixel value as a weighted average of other,

[*]Corresponding author
[**]Principal corresponding author
Email addresses: arian.maleki@rice.edu (Arian Maleki), manjari@rice.edu (Manjari Narayan), richb@rice.edu (Richard G. Baraniuk)
URL: http://dsp.rice.edu/ (Arian Maleki)
[1]Phone: +1 713.348.3579; Fax: +1 713.348.5685

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
26 NOV 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Anisotropic Nonlocal Means Denoising	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rice University,Department of Electrical and Computer Engineering,Houston,TX,77005**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

It has recently been proved that the popular nonlocal means (NLM) denoising algorithm does not optimally denoise images with sharp edges. Its weakness lies in the isotropic nature of the neighborhoods it uses to set its smoothing weights. In response, in this paper we introduce several theoretical and practical anisotropic nonlocal means (ANLM) algorithms and prove that they are near minimax optimal for edge-dominated images from the Horizon class. On real-world test images, an ANLM algorithm that adapts to the underlying image gradients outperforms NLM by a signi cant margin.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**47**	

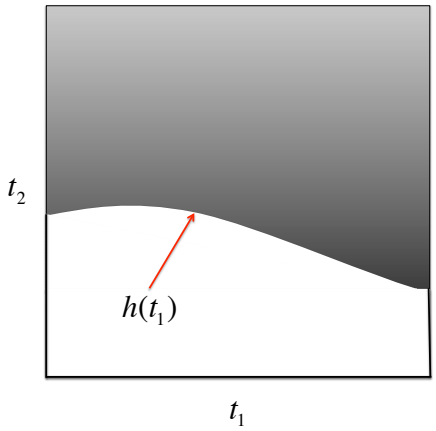


Figure 1: An example of a Horizon class image that features a smooth edge contour that separates the white region from the black region.

similar noisy pixels. However, instead of using spatial adjacency or noisy pixel value as the similarity measure to adjust the estimate weights, NLM uses a more reliable notion of similarity based on the resemblance of the pixels' neighborhoods in high-dimensional space. This unique feature benefits NLM in two ways. First, it provides more accurate weight estimates. Second, it enables NLM to exploit the contribution of all pixels in the image. In concert, these features enable NLM to provide state-of-the-art performance for a large class of image denoising problems.

Nevertheless, in a recent paper, we have proved that NLM does not attain optimal performance on images with sharp edges from the so-called *Horizon class* (see Figure 1) [8]. Indeed, NLM's theoretical performance is more or less equivalent to wavelet thresholding, which was shown to be suboptimal in [9]. The core problem is that NLM (and wavelet thresholding) cannot exploit the smoothness of the edge contour that separates the white and black regions. Therefore, there is room for improvement.

In this paper, we introduce and study a new denoising framework and prove that it is near-optimal for Horizon class images with sharp edges. *Anisotropic nonlocal means* (ANLM) outperforms NLM, wavelet thresholding, and more classical techniques by using anisotropic neighborhoods that are elongated along and matched to the local edge orientation. Figure 2 compares the neighborhoods used in ANLM with those used in NLM. Anisotropic

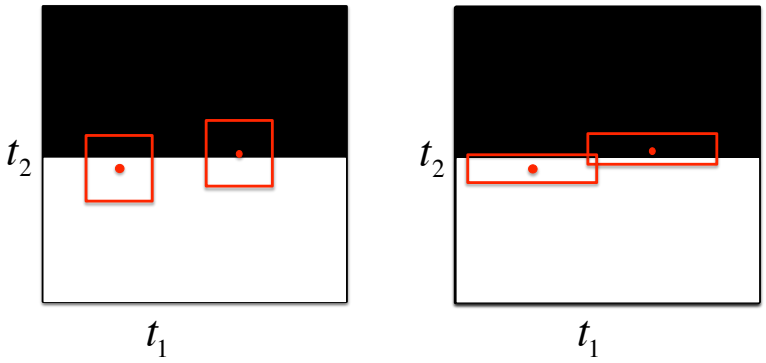


Figure 2: Comparison of (left) the isotropic neighborhoods employed by Non Local Means (NLM) versus (right) anisotropic neighborhoods employed by Anisotropic NLM (ANLM).

neighborhoods enable ANLM to distinguish between similar and dissimilar pixels more accurately.

We develop three different ANLM algorithms of increasing levels of practicality. *Oracle Anisotropic Nonlocal Means* (OANLM) assumes perfect knowledge of the local orientation of the edge contour and is used primarily for our theoretical optimality analysis. *Discrete-angle Anisotropic Nonlocal Means* (DANLM) optimizes the choice of the anisotropic neighborhood around each pixel in order to achieve near-optimal performance without any oracle information. Since it is more computationally demanding than NLM, we introduce an algorithmic simplification. *Gradient based Anisotropic Nonlocal Means* (GANLM) uses image gradient information to estimate the edge orientation; using simulations, we demonstrate that GANLM significantly outperforms NLM in practice on both Horizon class and real-world images.

The paper is organized as follows. Section 2 explains the minimax framework we use to analyze the denoising algorithms and reviews the necessary background. Section 3 introduces the OANLM and DANLM algorithms and presents the main theorems. Section 4 addresses some of the practical ANLM issues by introducing GANLM and summarizes the results of a range of simulations using synthetic and real-world imagery. Section 5 contains the proofs of the main theorems. Section 6 reviews the related work in the literature. Section 7 closes with a discussion of our current and potential future results.

2. Minimax analysis framework

In this section we introduce the minimax framework [2, 10] and the Horizon class image model considered in this paper. Note that, in order to streamline the proofs, we take a continuous-variable analysis approach in this paper, in contrast to our approach in [8]. The moral of the story is compatible with [8], however.

2.1. Risk

We are interested in estimating an image described by the function $f : [0,1]^2 \to [0,1]$ ($f \in L^2([0,1]^2)$) from its noisy observation

$$dY(t_1,t_2) = f(t_1,t_2)dt_1dt_2 + \sigma dW(t_1,t_2). \tag{1}$$

Without loss of generality, we consider only square images. Here $W(t_1,t_2)$ is the Wiener sheet,[2] and σ is a constant that scales the noise. For a given function f and a given estimator $\hat{f}$, define the *risk* as

$$R(f,\hat{f}) = \mathbb{E}(\|f-\hat{f}\|_2^2),$$

where the expected value is over W. The risk can be decomposed into bias squared and variance terms

$$R(f,\hat{f}) = \|f-\mathbb{E}\hat{f}\|_2^2 + \mathbb{E}\|\hat{f}-\mathbb{E}\hat{f}\|_2^2.$$

Let f belong to a class of functions $\mathcal{F}$. The risk of the estimator $\hat{f}$ on $\mathcal{F}$ is defined as the risk of the least-favorable function, i.e.,

$$R(\mathcal{F},\hat{f}) = \sup_{f\in\mathcal{F}} R(f,\hat{f}).$$

The *minimax* risk over the class of functions $\mathcal{F}$ is then defined as

$$R^*(\mathcal{F}) = \inf_{\hat{f}} \sup_{f\in\mathcal{F}} R(f,\hat{f}).$$

$R^*(\mathcal{F})$ is a lower bound for the performance of any estimator on $\mathcal{F}$.

In this paper, we are interested in the asymptotic setting as $\sigma \to 0$. For all the estimators we consider, $R(\mathcal{F},\hat{f}) \to 0$ as $\sigma \to 0$. Therefore, following [11, 12] we will consider the *decay rate of the minimax risk* as our measure of performance.

[2] The Wiener sheet is the primitive of white noise.

2.2. Horizon edge model

In our analysis, we consider the *Horizon* model that contains piecewise constant images with sharp step edges lying along a smooth contour [9, 11, 13] (our analysis extends easily to piecewise smooth edges). Let $H\ddot{o}lder^{\alpha}(C)$ be the class of Hölder functions on $\mathbb{R}$, defined in the following way: $h \in H\ddot{o}lder^{\alpha}(C)$ if and only if

$$|h^{(k)}(t_1) - h^{(k)}(t_1')| \leq C|t_1 - t_1'|^{\alpha-k},$$

where $k = \lfloor \alpha \rfloor$. Consider a transformation that maps each one-dimensional edge contour function h to a two-dimensional image $f_h : [0,1]^2 \to \mathbb{R}$ via

$$f_h(t_1, t_2) = \mathbf{1}_{\{t_2 < h(t_1)\}}.$$

The Horizon class of images is then defined as

$$H^{\alpha}(C) = \{f_h(t_1, t_2) : h \in H\ddot{o}lder^{\alpha}(C) \cap H\ddot{o}lder^{1}(1)\}, \tag{2}$$

where α is the smoothness of the edge contour. Figure 1 illustrates a sample function of this class. The following theorem characterizes the minimax rate of $H^{\alpha}(C)$ [9, 11, 13].[3]

Theorem 1. [9, 11, 13] *For $\alpha \geq 1$, the minimax risk of the class $H^{\alpha}(C)$ is*

$$R^*(H^{\alpha}(C)) = \Theta\left(\sigma^{\frac{2\alpha}{\alpha+1}}\right).$$

Achieving the minimax rate is a laudable goal that any well-respecting denoising algorithm should aspire to. In this paper, we will focus primarily on $\alpha = 2$ edge contours, for which the optimal minimax decay rate is $\sigma^{-4/3}$. However, it is straightforward to draw similar conclusions for the other values of α.

[3]The models considered in [9, 11, 13] are slightly different from the continuous framework of this paper. Therefore, for the sake of completeness we prove Theorem 1 in Appendix C.

Table 1: Minimax risk decay rates of several classical image denoising algorithms; recall that the optimal minimax rate is $\sigma^{4/3}$.

Algorithm	Minimax Rate
Mean filter [14]	$\sigma^{2/3}$
Wavelet thresholding [9, 11]	σ^{1}
Nonlocal means [8]	σ^{1}

2.3. Image denoising algorithms

Minimax risk analysis of the classical denoising algorithms has revealed their suboptimal performance on images with sharp edges. Table 1 summarizes several of these algorithms and their minimax decay rates (up to a log factor).[4] The suboptimality of wavelet thresholding has led to the development of ridgelets [15], curvelets [16], wedgelets [9], platelets [17], shearlets [18], contourlets [19], bandelets [20], directionlets [21] and other types of directional transforms. See [22] and the references therein for more information. In the framework of this paper, wedgelet denoising [9] provably achieves the optimal minimax rate. However, it performs poorly on textures, which has limited its application in image processing.

2.4. Nonlocal means denoising

In 2005, Buades, Coll, and Morel significantly improved the performance of the bilateral filter [8] by incorporating a new notion of pixel similarity. As in the bilateral filter, NLM estimates each pixel value using a weighted average of other pixel values in the image. However, NLM sets the weights according to the similarity between the pixel neighborhoods rather than the pixel values. Furthermore, in contrast to the bilateral filter, in which only the vicinity of each pixel contributes to the estimate, in NLM all pixels may contribute.

[4] The analysis framework used in [14] and [8] is discrete rather than continuous. Therefore, for the sake of completeness we establish the results for the mean filter and NLM in Appendix A and Appendix B, respectively.

Specifically, the NLM algorithm is defined as follows. Let $S = [0,1]^2$ represent the domain of the image. Define the δ-neighborhood of a (t_1, t_2) as

$$I_\delta(t_1,t_2) = \left[t_1 - \frac{\delta}{2}, t_1 + \frac{\delta}{2}\right] \times \left[t_2 - \frac{\delta}{2}, t_2 + \frac{\delta}{2}\right]. \tag{3}$$

Following the definition of NLM in the discrete setting, we pixelate this neighborhood by partitioning $I_\delta(t_1,t_2)$ into n^2 subregions $I_\delta^{j_1,j_2} = [t_1 - \frac{j_1\delta}{2n}, t_1 + \frac{j_1\delta}{2n}] \times [t_2 - \frac{j_2\delta}{2n}, t_2 + \frac{j_2\delta}{2n}]$. The pixelated neighborhood of (t_1,t_2) is defined as $\mathbf{y}^\delta_{t_1,t_2} \in \mathbb{R}^{n\times n}$ and satisfies

$$\mathbf{y}^\delta_{t_1,t_2}(j_1,j_2) = \int_{(s_1,s_2)\in I_\delta^{j_1,j_2}} dY(s_1,s_2).$$

We further define the pixelated process as

$$X(t_1,t_2) = \frac{n^2}{\delta^2} \int_{(s_1,s_2)\in I_{\delta/n}(t_1,t_2)} dY(s_1,s_2).$$

Define the δ-neighborhood distance between two points in the image as

$$\begin{aligned} &d^2_\delta(dY(t_1,t_2), dY(s_1,s_2)) \\ &\quad = \frac{1}{n^2-1}\left(\|\mathbf{y}^\delta_{t_1,t_2} - \mathbf{y}^\delta_{s_1,s_2}\|_2^2 - |\mathbf{y}^\delta_{t_1,t_2}(0,0) - \mathbf{y}^\delta_{s_1,s_2}(0,0)|^2\right), \end{aligned} \tag{4}$$

where $\|\mathbf{y}^\delta_{t_1,t_2}\|_2^2 \triangleq \sum_{i,j} \mathbf{y}^\delta_{t_1,t_2}(i,j)$. Note that in contrast to the definition in [7], we have removed the center element $|\mathbf{y}^\delta_{t_1,t_2}(0,0) - \mathbf{y}^\delta_{s_1,s_2}(0,0)|^2$ from the summation. Since we assume that $n^2 \to \infty$ as $\sigma \to 0$, the effect is negligible on the asymptotic performance. But, as we will see below, removing the center element simplifies the calculations considerably. NLM uses the neighborhood distances to estimate

$$\hat{f}^N(t_1,t_2) = \frac{\int_{(s_1,s_2)\in S} w^N_{t_1,t_2}(s_1,s_2) X(s_1,s_2) ds_1 ds_2}{\int_{(s_1,s_2)\in S} w^N_{t_1,t_2}(s_1,s_2) ds_2 ds_2},$$

where $w^N_{t_1,t_2}(s_1,s_2)$ is set according to the δ-neighborhood distance between $dY(t_1,t_2)$ and $dY(s_1,s_2)$.[5] It is straightforward to verify that $\mathbb{E}d^2_\delta(dY(t_1,t_2),$

[5]We assume that both $w^N_{t_1,t_2}(s_1,s_2)X(s_1,s_2)$ and $w^N_{t_1,t_2}(s_1,s_2)$ are Lebesgue integrable with high probability.

$dY(s_1, s_2)) = d_\delta^2(f(t_1, t_2), f(s_1, s_2)) + \frac{2n^2\sigma^2}{\delta^2}$, which suggests the following strategy for setting the weights:

$$w_{t_1,t_2}^N(s_1, s_2) = \begin{cases} 1 & \text{if } d_\delta^2(dY(t_1, t_2), dY(s_1, s_2)) \leq \frac{2n^2\sigma^2}{\delta^2} + \tau_\sigma, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where τ_σ is the *threshold parameter*. Soft/tapered weights have been explored and are often used in practice [7]. However, the above untapered weights capture the essence of the algorithm while simplifying the analysis.

The distinguishing feature of NLM — the weighted averaging of pixels based on the neighborhoods — produces a decay rate that is superior to that linear filters as shown in [8]. However, compared to the optimal rate of $\sigma^{4/3}$, NLM remains suboptimal. We introduce the anisotropic NLM algorithm in the next section to address this gap in performance.

3. Anisotropic nonlocal means denoising

In this section, we introduce the Anisotropic Nonlocal Means (ANLM) algorithm concept that exploits the smoothness of edge contours via anisotropic neighborhoods. We then present OANLM and DANLM algorithms and explain their performance guarantees.

3.1. Directional neighborhoods

We now formally introduce the notion of directional neighborhoods. Then we extend NLM to exploit such neighborhoods. This will lay the foundation for the OANLM and DANLM algorithms.

Let $R_{\nu,\mu}^\theta(\cdot)$ represent the rotation operator on a neighborhood; when applied to a generic point $(u, v) \in S$ in the neighborhood around (ν, μ), $R_{\nu,\mu}^\theta$ rotates (u, v) by θ° counter-clockwise around the point (ν, μ). For a set $Q \subset S$ we define $Q_\theta \triangleq R_{\nu,\mu}^\theta(Q)$ as

$$(s, t) \in Q_\theta \Leftrightarrow \exists (u, v) \in Q \text{ such that } (s, t) = R_{\nu,\mu}^\theta((u, v)).$$

The $(\theta, \delta_s, \delta_\ell)$-*anisotropic neighborhood* of the point (t_1, t_2) is defined as

$$I_{\theta,\delta_s,\delta_\ell}(t_1, t_2) \triangleq R_{t_1,t_2}^\theta \left(\left[t_1 - \frac{\delta_\ell}{2}, t_1 + \frac{\delta_\ell}{2} \right] \times \left[t_2 - \frac{\delta_s}{2}, t_2 + \frac{\delta_s}{2} \right] \right) \cap S.$$

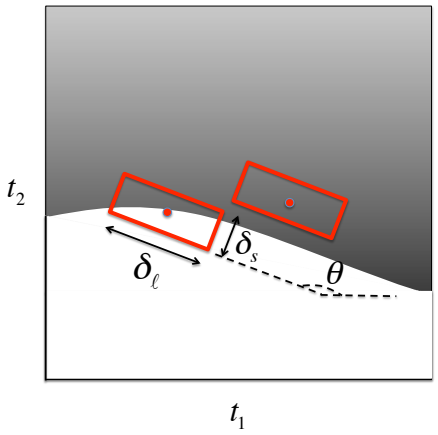


Figure 3: The anisotropic neighborhood $I_{\theta,\delta_s,\delta_\ell}(t_1,t_2)$ in the discrete setting for two different pixels of a Horizon class image.

where θ, δ_ℓ, and δ_s ($\delta_s \leq \delta_\ell$) represent the orientation angle, length, and width of the neighborhood, respectively. Figure 3 displays such neighborhoods for two different pixels. Each element of the NLM partitioning of $I_{\theta,\delta_s,\delta_\ell}(t_1,t_2)$ into $n_s \times n_\ell$ pixelated regions has a corresponding element in the partitioning of neighborhoods, given by $I^{j_1,j_2}_{\theta,\delta_s,\delta_\ell}(t_1,t_2) = R^{\theta}_{t_1,t_2}([t_1 - \frac{j_1\delta_\ell}{2n_\ell}, t_1 + \frac{j_1\delta_\ell}{2n_\ell}] \times [t_1 - \frac{j_1\delta_s}{2n_s}, t_1 + \frac{j_1\delta_s}{2n_s}])$.

The pixelated neighborhood of (t_1,t_2) is defined as the $y \in \mathbb{R}^{n_s \times n_\ell}$ satisfying

$$\mathbf{y}^{\theta,\delta_\ell,\delta_s}_{t_1,t_2}(j_1,j_2) = \frac{n_s n_\ell}{\delta_s \delta_\ell} \int\int_{(s_1,s_2)\in I^{j_1,j_2}_{\theta,\delta_s,\delta_\ell}(t_1,t_2)} dY(s_1,s_2).$$

Define the neighborhood process as

$$X(t_1,t_2) = \frac{n_s n_\ell}{\delta_s \delta_\ell} \int\int_{(s_1,s_2)\in I_{\theta,\frac{\delta_s}{n_s},\frac{\delta_\ell}{n_\ell}}(t_1,t_2)} dY(s_1,s_2).$$

The anisotropic $(\delta_\ell,\delta_s,\theta)$-neighborhood distance between two given points in the image is then defined as

$$\begin{aligned} &d^2_{\theta,\delta_s,\delta_\ell}(dY(t_1,t_2), dY(s_1,s_2)) \\ &\quad = \frac{1}{n_s n_\ell - 1}\left(\|\mathbf{y}^{\theta,\delta_\ell,\delta_s}_{t_1,t_2} - \mathbf{y}^{\theta,\delta_\ell,\delta_s}_{s_1,s_2}\|_2^2 - |\mathbf{y}^{\theta,\delta_\ell,\delta_s}_{t_1,t_2}(0,0) - \mathbf{y}^{\theta,\delta_\ell,\delta_s}_{s_1,s_2}(0,0)|^2\right). \quad (6) \end{aligned}$$

The ANLM estimate at point (t_1, t_2) is given by

$$\hat{f}^{\theta,\delta_s,\delta_\ell}(t_1,t_2) = \frac{\int_{(s_1,s_2)\in S} w^{\theta,\delta_s,\delta_\ell}_{t_1,t_2}(s_1,s_2)X(s_1,s_2)ds_1ds_2}{\int\int w^{\theta,\delta_s,\delta_\ell}_{t_1,t_2}(s_1,s_2)ds_1ds_2}.$$

where the weights are obtained from

$$w^{\theta,\delta_s,\delta_\ell}_{t_1,t_2}(s_1,s_2) = \begin{cases} 1 & \text{if } d^2_{\theta,\delta_s,\delta_\ell}(dY(t_1,t_2), dY(s_1,s_2)) \leq \frac{2n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Here τ_σ is the *threshold parameter* and $\frac{2n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma$ is the *threshold value*.

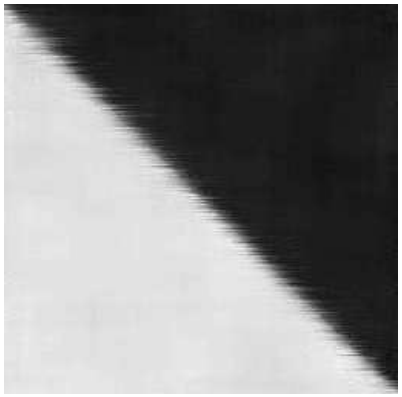
ANLM extends NLM in two significant ways. First, in ANLM $\theta, \delta_s, \delta_\ell$ are free parameters, while in NLM $\theta = 0$ and $\delta_s = \delta_\ell$. As Figures 4 and 5 demonstrate, these free parameters significantly affect the performance of ANLM. Figure 4 illustrates the visual denoising results as we vary the angle of the anisotropic neighborhood. Figure 5 does the same for the empirical peak signal-to-noise ratio (PSNR).[6]

3.2. Oracle ANLM

It is clear from Figures 4 and 5 that we should align the ANLM neighborhood locally with the edge to maximize denoising performance. In the Oracle ANLM (OANLM) algorithm, we assume that we have access to an oracle that knows the image's edge orientation at each pixel. Consider $f_h(t_1, t_2) \in H^\alpha(C)$. Let $h'(t_1)$ denote the derivative of the edge contour $h(t_1)$, and let $\Gamma_\gamma \triangleq \{(t_1, t_2) : t_1 = \gamma\}$ segment the region S. Define θ_γ as the angle between the tangent to the edge contour h and the horizontal line at $t_1 = \gamma$. OANLM is defined as ANLM with the following parameter settings:

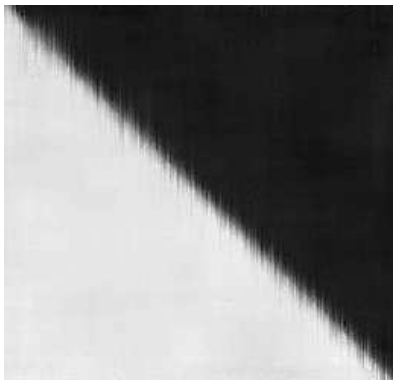
- *Quadratic scaling:* Instead of using isotropic (square) neighborhoods, we use anisotropic (rectangular) neighborhoods of size $\delta_s \times \delta_\ell$. The scaling of δ_s and δ_ℓ depends on the smoothness of the edge. Since

[6] In practice images are measured on a discrete lattice, i.e., the pixelated values are known. Let the pixelated values of the original noise-free image and the estimator be f_{ij} and $\hat{f}_{i,j}$, respectively. The empirical mean-squared-error (MSE) of an estimator $\hat{f}$ is defined as $\text{MSE} = \frac{1}{n^2}\sum_{i,j}(\hat{f}_{i,j} - f_{i,j})^2$. If $f_{i,j} \in [0, A]$, then the empirical peak signal-to-noise ratio (PSNR) is defined as $10\log_{10}\frac{A^2}{\text{MSE}}$.



(a) 0°, PSNR = 20.6dB

(b) 45°, PSNR = 19.2dB



(a) 90°, PSNR = 20.6dB

(d) 135°, PSNR = 28.9dB

Figure 4: Effect of the anisotropic neighborhood orientation angle θ on the denoising performance of ANLM. The images are 256×256 pixels with a true edge orientation of 135°, and the noise standard deviation in practical algorithms is $\xi = 0.5$ (see Section 4.1). The neighborhood size and threshold parameter are $\delta_s = 1/256$, $\delta_\ell = 20/256$ and $\tau = 2.7\xi^2$, respectively. The angles of the neighborhoods (θ) and the resulting PSNR of the estimates are noted below the images. There is a substantial subjective and objective improvement in the performance of ANLM when the neighborhood is aligned with the edge.

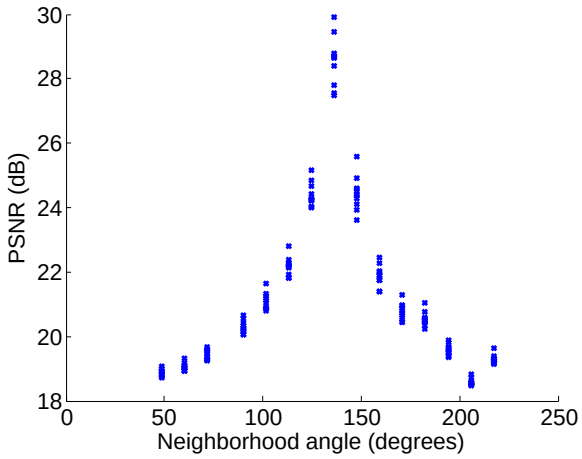


Figure 5: Effect of the anisotropic neighborhood orientation angle θ on the denoising performance of ANLM. The experimental parameters are the same as in Figure 4. For each value of θ we plot the result of 10 Monte Carlo simulations.

we are primarily interested in $H^2(C)$, we will emphasize the quadratic scaling[7] $\delta_s = 4\sigma^{4/3}|\log\sigma|^{4/3}$ and $\delta_\ell = 2\sigma^{2/3}|\log\sigma|^{2/3}$; that is $\delta_s = \delta_\ell^2$.

- *Aligned neighborhoods:* We align the neighborhood at point (t_1, t_2) to the orientation θ_{t_1} of the edge contour at $(t_1, h(t_1))$. Thus all pixels in a column have the same orientation. Figure 6 illustrates such a neighborhood selection.[8]

- *Logarithmic pixelation:* We assume that $n_s \times n_\ell = \log^2\sigma$. In other words, as σ increases the resolution of the pixelated image increases as well.

Theorem 2. *If $\hat{f}^O$ is the OANLM estimator with the threshold $\tau_\sigma = \frac{2}{\sqrt{|\log\sigma|}}$,*

[7]If the smoothness of an edge is $\alpha > 1$, then the optimal neighborhood size is given by $\delta_s = \Theta(\sigma^{-2\alpha/(\alpha+1)}|\log\sigma|^{2\alpha/(\alpha+1)})$ and $\delta_\ell = \Theta(\sigma^{-2/(\alpha+1)}|\log\sigma|^{2/(\alpha+1)})$.

[8]In smooth regions, where the pixel neighborhoods do not intersect with the edge, neither the shape nor the orientation of the neighborhood affect the denoising performance.

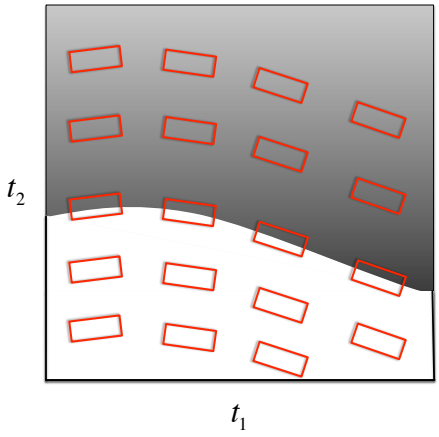


Figure 6: The OANLM neighborhoods satisfy the quadratic scaling $\delta_s = \delta_\ell^2$ and are aligned with the edge. In regions far from the edge, isotropic neighborhoods perform just as well.

then

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma^{4/3}|\log\sigma|^{4/3}).$$

The proof of Theorem 2 can be simply modified to provide a bound for the risk of NLM as well. Let $\delta = \sqrt{\delta_s \delta_\ell}$, where δ_s and δ_ℓ are as given in OANLM algorithm.

Corollary 1. *If $\hat{f}^N$ is the NLM estimator with the threshold $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$, then*

$$R(H^\alpha(C), \hat{f}^N) = O(\sigma|\log\sigma|).$$

In fact, under certain mild assumptions, the risk of NLM can be lower bounded by $\Omega(\sigma)$. The proof is similar to the proof of Theorem 5 in [8]. However, for the sake of completeness and since our framework differs from [8], we overview the main steps of the proof in Appendix B.

Theorem 2 is based on the strong oracle assumption that the edge direction is known exactly. Needless to say, such information is rarely available in

practical applications. Consider a weaker notion of OANLM that has access to an estimate $\hat{\theta}_\gamma$ of θ_γ that satisfies

$$|\hat{\theta}_\gamma - \theta_\gamma| \leq \Theta(\sigma^\beta). \tag{8}$$

OANLM with exact edge orientation information corresponds to $\beta = \infty$ in this model. When $\beta < \infty$, the choices $\delta_s = \sigma^{2/3}|\log\sigma|^{2/3}$ and $\delta_\ell = \sigma^{4/3}|\log\sigma|^{4/3}$ are not necessarily optimal. The following theorem characterizes the performance of OANLM using $\hat{\theta}$.

Theorem 3. *Let the OANLM estimator use the inaccurate edge orientation $\hat{\theta}_\gamma$ satisfying* (8)*. Set the neighborhood sizes to $\delta_\ell = \min(\sigma^{2/3}|\log^{2/3}\sigma|, \sigma^{-1+\beta/2}|\log\sigma|)$ and $\delta_s = \sigma^2\log^2\sigma/\delta_\ell$. Then the risk of the estimator with $\tau_\sigma = 2/\sqrt{|\log(\sigma)|}$ satisfies*

$$R(H^\alpha(C), \hat{f}^O) = O(\delta_s \mathrm{Poly}(|\log\sigma|)),$$

where $\mathrm{Poly}(|\log\sigma|)$ *is a polynomial of degree at most* 2 *in terms of* $|\log\sigma|$*.*

If the edge estimate is exact, then the result simplifies to the result of Theorem 2. However, this theorem confirms that OANLM algorithm achieves the optimal rate even if there is an error in $\hat{\theta}$ of order $O(\sigma^\beta)$ with $\beta > 2/3$.

Corollary 2. *Let the OANLM estimator use the inaccurate edge orientation $\hat{\theta}_\gamma$ satisfying* (8)*. Set the neighborhood sizes to $\delta_\ell = \sigma^{1-\beta/2}|\log\sigma|$ and $\delta_s = \sigma^{1+\beta/2}|\log\sigma|$. Then, the risk of the estimator satisfies*

$$R(H^\alpha(C), \hat{f}^O) = O(\sigma^{1+\beta/2}\mathrm{Poly}(|\log\sigma|)).$$

This corollary suggests that, as long as we have an edge orientation estimate that improves as the number of pixels increases, OANLM outperforms NLM. Also note that, as β decreases, the neighborhoods become more isotropic.

3.3. Discrete angle ANLM

In this section, we introduce the Discrete-angle ANLM (DANLM) algorithm that achieves the minimax rate without oracle information. The key

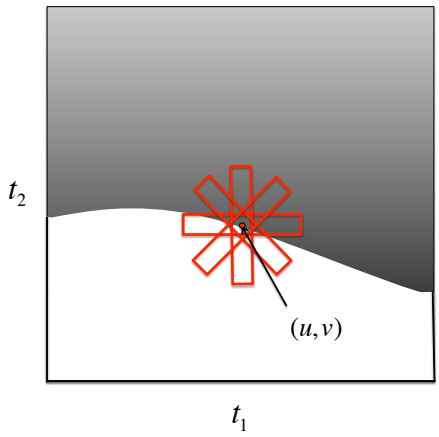


Figure 7: DANLM neighborhoods around one pixel location for $q = 4$.

idea is to calculate the neighborhood distance over several directional neighborhoods and fuse them to obtain a similarity measure that works well for all directions. As for OANLM, set $\delta_s = \sigma^{4/3}|\log(\sigma)|^{4/3}$ and $\delta_\ell = \sigma^{2/3}|\log(\sigma)|^{2/3}$, and let $q = \pi\sigma^{-2/3}$. Define the angles $\theta_0 = 0$, $\theta_1 = \sigma^{2/3}$, $\theta_2 = 2\sigma^{2/3}, \ldots, \theta_q = \pi - \sigma^{2/3}$. For a point (t_1, t_2) we consider all of the anisotropic, directional neighborhoods for $\theta \in \Theta$. Figure 7 displays four of these neighborhoods.

Define the discrete angle anisotropic distance between $dY(t_1, t_2)$ and $dY(s_1, s_2)$

$$d^2_{\mathcal{A}}(dY(t_1,t_2), dY(s_1,s_2)) \triangleq \min_{\theta\in\Theta} d^2_{\theta,\delta_s,\delta_\ell}(dY(t_1,t_2), dY(s_1,s_2)),$$

where $d^2_{\theta,\delta_s,\delta_\ell}(dY(t_1,t_2), dY(s_1,s_2))$ is defined in (6). The DANLM estimate is then given by

$$\hat{f}^D(t_1,t_2) = \frac{\int_{(s_1,s_2)\in S} w^{AN}_{t_1,t_2}(s_1,s_2)X(s_1,s_2)ds_1ds_2}{\int_{(s_1,s_2)\in S} w^{AN}_{t_1,t_2}(s_1,s_2)ds_1ds_2}.$$

where

$$w^{AN}_{t_1,t_2}(s_1,s_2) = \begin{cases} 1 & \text{if } d^2_{\mathcal{A}}(dY(t_1,t_2), dY(s_1,s_2)) \le \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} + \tau_\sigma, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

In summary, we note the following features of the DANLM algorithm. First, it uses the quadratic scaling $\delta_s = \delta_\ell^2$. Second, the optimal neighborhood

direction can change from pixel to pixel. The following theorem, proved in Section 5, shows that the risk of this algorithm is within the logarithmic factor of the minimax risk.

Theorem 4. *The risk of DANLM satisfies*

$$R(H^\alpha(C_\alpha), \hat{f}^D) \leq O(\sigma^{4/3} |\log(\sigma)|^{4/3}).$$

Figure 8 reprises the simulation experiment of Figure 4 using the four algorithms described thus far: NLM, OANLM with perfect knowledge of the edge orientation, OANLM with imperfect knowledge of the edge orientation, and DANLM with four angles. All three of the latter algorithms outperform the isotropic NLM.

4. Practical ANLM algorithms and experiments

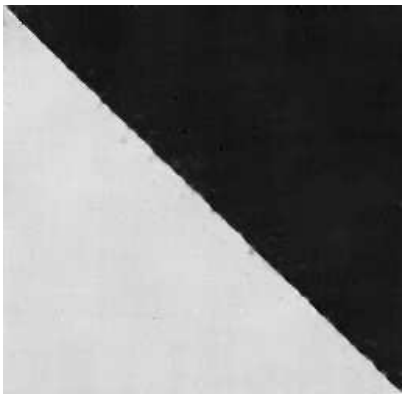
In this section we introduce a practical gradient-based ANLM algorithm and then complement the above theoretical arguments with additional simulations and experiments with synthetic and real-world imagery.

4.1. Extension to discrete images

In practice the observations are noisy pixelated values of an image and the objective is only to estimate the pixelated values. In this section we explain how the ideas of directional neighborhood and ANLM can be extended to the discrete settings. Suppose we are interested in estimating a $n \times n$ image $f(\frac{i}{n}, \frac{j}{n})$ with noisy observations $o_{i,j} = f(\frac{i}{n}, \frac{j}{n}) + \zeta_{i,j}$, where $\zeta_{i,j} \overset{iid}{\sim} N(0, \xi^2)$ and ξ is the standard deviation of the noise. The extension of the anisotropic neighborhood to the discrete setting is straightforward. Let $\bar{S} = \left\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\} \times \left\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}$ and $\tilde{S} = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. For a given set $B \subset S$ we define $\bar{B} = B \cap \bar{S}$.

The discrete $(\theta, \delta_s, \delta_\ell)$-distance between two pixels $o_{i,j}$ and $o_{m,\ell}$ is defined as

$$\bar{d}^2_{\theta,\delta_s,\delta_\ell}(o_{i,j}, o_{m,\ell}) \triangleq \frac{1}{|\mathcal{P}|} \sum_{(r,q)\in\mathcal{P}} (o_{i+r,j+q} - o_{m+r,\ell+q})^2, \tag{10}$$



(a) NLM, PSNR = 26 dB

(b) OANLM with no error, PSNR = 28.9 dB

(c) ONALM with error, PSNR = 28.6 dB

(d) DANLM, PSNR = 27.5 dB

Figure 8: Continuation of the experiment of Figure 4 that compares (a) isotropic NLM, (b) OANLM with perfect knowledge of the edge direction, (c) OANLM with 10% error in the knowledge of the edge direction, (d) DANLM with $q = 4$ angles. The images are of size 256×256 pixels. In the ANLM algorithms, δ_s and δ_ℓ are the same as in Figure 4. The neighborhood size of NLM is $\sqrt{\delta_s \times \delta_\ell}$. The edge orientation is 135°. The standard deviation of the noise is $\xi = 0.5$.

where $\mathcal{P} = \{(r, q) \in \mathbb{Z}^2 \mid (\frac{i+r}{n}, \frac{j+q}{n}) \in \bar{I}_{\theta,\delta_s,\delta_\ell}(i/n, j/n)\}$. See Figure 9. The ANLM estimate at pixel $f(\frac{i}{n}, \frac{j}{n})$ is given by

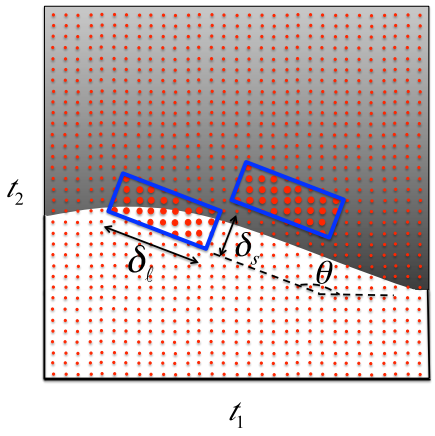


Figure 9: The anisotropic neighborhood $\bar{I}_{\theta,\delta_s,\delta_\ell}(\frac{i}{n},\frac{j}{n})$ in the discrete setting for two different pixels of a very simple Horizon class image.

$$\hat{f}_{i,j}^{\theta,\delta_s,\delta_\ell} = \frac{\sum_m \sum_\ell \bar{w}_{i,j}^{\theta,\delta_\ell,\delta_s}(m,\ell) o_{m,\ell}}{\sum_m \sum_\ell \bar{w}_{i,j}^{\theta,\delta_\ell,\delta_s}(m,\ell)},$$

where the weights are obtained from the distances $\bar{d}^2_{\theta,\delta_s,\delta_\ell}(o_{i,j}, o_{m,\ell})$. For instance, the simple policy introduced in (7) corresponds to

$$\bar{w}_{i,j}^{\theta,\delta_\ell,\delta_s}(m,\ell) = \begin{cases} 1 & \text{if } \bar{d}^2_{\theta,\delta_s,\delta_\ell}(o_{i,j}, o_{m,\ell}) \le 2\xi^2 + \tau_n, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Using the discrete anisotropic neighborhood and distance, the extensions of OANLM and DANLM to the discrete setting is straightforward.

4.2. Gradient-based ANLM

While DANLM is somewhat practical and also theoretically optimal for the Horizon class of images, its computational complexity is higher than NLM and grows linearly in the number of directions q, where $q \sim o(n^{2/3})$. As a more practical alternative, we propose Gradient based ANLM (GANLM), which adjusts the orientation of an anisotropic ANLM neighborhood using an estimate of the edge orientation provided by the image gradients. Pseudocode is given in Algorithm 4.2. Note that GANLM reverts back to NLM in regions with low image gradients, since they will not be "edgy" enough to benefit from the special treatment.

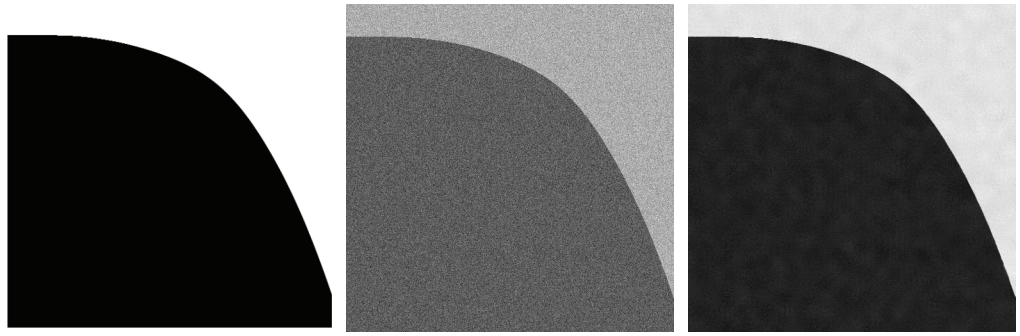
Algorithm 1 *Gradient-based ANLM (GANLM)*

Inputs:
$\hat{f}_{i,j}$: Estimate of the image pixel
$\delta_s \times \delta_\ell$: Size of the neighborhood
λ: Threshold that determines isotropic/anisotropic neighborhood selection

Estimate image gradient $(g_h(i,j), g_v(i,j))$ at each pixel (i,j)
for every pixel $(i,j) \in I$ **do**
  $g(i,j) = \sqrt{g_h^2(i,j) + g_v^2(i,j)}$
  $\theta_{i,j} = \tan^{-1}\left(\dfrac{g_v(i,j)}{g_h(i,j)}\right)$
  if $g_i \geq \lambda$ **then**
    Perform ANLM at pixel $y_{i,j}$ with $d_{\delta_\ell,\delta_s,\theta_{i,j}}$
  else
    Perform NLM at pixel $y_{i,j}$ with $\sqrt{\delta_s \delta_\ell}$.
  end if
end for

There is a rich literature on robust image gradient estimation [23–25]. Most simply, if $g_h(i,j)$ and $g_v(i,j)$ are the estimated image derivatives at pixel (i,j), then we can estimate the local orientation of an edge by $\hat{\theta}(i,j) = \tan^{-1}\left(\frac{g_v(i,j)}{g_h(i,j)}\right)$. To allay any concerns that gradient-based adaptivity is not robust to noise and errors, we recall Theorem 3, which establishes the robustness of OANLM to edge angle estimation error. For extremely noisy images, numerous heuristics are possible, including estimating the image gradients for GANLM from an isotropic NLM pilot estimate. Figure 10 builds on Figure 8 with a slightly more realistic, curved edge and demonstrates that the pilot estimate approach to GANLM performs almost as well as an oracle GANLM that has access to the edge gradient.

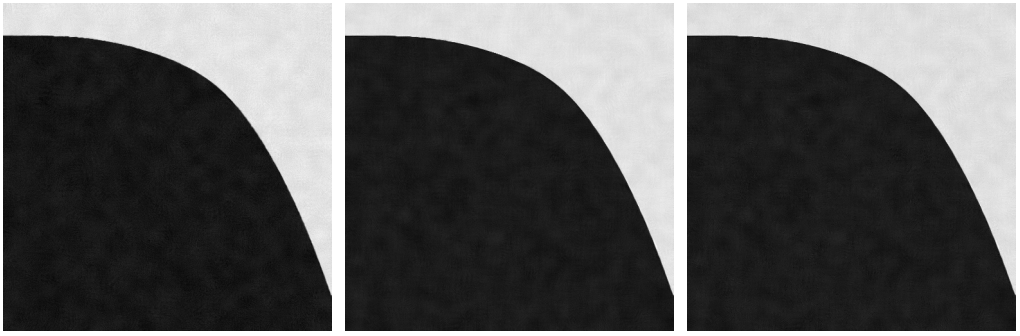
Table 2 summarizes the performance of the algorithms introduced in this paper with that of NLM on the natural test images Barbara [26], Boats [27], and Wet Paint [28] submerged in noise of two different levels. The table demonstrates two facts. First, the performance of the practical GANLM algorithm is very close to the oracle GANLM algorithm. Second these two algorithms outperform DANLM in all but one case and significantly out-



(a) Horizon image

(b) Noisy image, PSNR = 16.5dB

(c) NLM, PSNR = 34.4dB



(d) DANLM, PSNR = 36.2dB

(e) Oracle GANLM, PSNR = 38dB

(f) GANLM, PSNR = 38dB

Figure 10: Simulation experiment in the same vein as Figures 4 and 8 but with a slightly more realistic C^2 curved edge. The images are of size 512×512 pixels. In the ANLM algorithms δ_s and δ_ℓ follow the quadratic scaling of Theorem 2. The neighborhood size of NLM is $\sqrt{\delta_s \times \delta_\ell}$. The standard deviation of the noise is $\xi = 0.15$.

Table 2: Performance of NLM and ANLM algorithms on three test images at two noise levels.

Test image	Algorithm	$\xi = 0.25$	$\xi = 0.15$
Barbara (512 x 512)	NLM	22.48	25.86
	DANLM	23.15	26.27
	Oracle GANLM	23.51	26.63
	GANLM	23.50	26.60
Boats (512 x 512)	NLM	22.75	25.83
	DANLM	23.45	26.45
	Oracle GANLM	23.88	26.45
	GANLM	23.75	26.35
Wet Paint (1024 x 1024)	NLM	27.66	30.49
	DANLM	28.41	30.75
	Oracle GANLM	29.02	31.18
	GANLM	28.86	31.07

perform standard NLM in all cases. We use 4 discrete angles[9] in DANLM and attribute the superior performance of GANLM over DANLM to more accurate selection of orientations.

5. Proofs of the main results

5.1. Preamble

We first introduce some notation. Define the following partitions of the set S:

$$\begin{aligned} S_1 &\triangleq \{(v,u) \mid u > h(v) + (1 + C/2)\delta_s\}, \\ S_2 &\triangleq \{(v,u) \mid h(v) - (1 + C/2)\delta_s < u < h(v) + (1 + C/2)\delta_s\}, \\ S_3 &\triangleq \{(v,u) \mid u < h(v) - (1 + C/2)\delta_s\}. \end{aligned}$$

It is important to note that, if $(t_1, t_2) \in S_1$ and $\tan(\theta) = h'(t_1)$, then $I_{\theta,\delta_s,\delta_\ell}(t_1, t_2)$ does not overlap with the edge contour. In other words, the correctly aligned neighborhood of $(t_1, t_2) \in S_1$ is always above the edge. The points in S_3 satisfy a similar property. This is clarified in Figure 11.

[9]Experiments with larger q did not differ appreciably in PSNR at this image size.

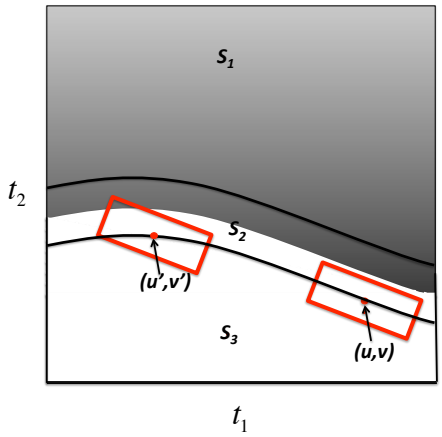


Figure 11: Regions S_1, S_2 and S_3. The neighborhood of $(u, v) \in S_3$ is aligned with the edge contour, and therefore it does not intersect the edge, while the neighborhood of (u', v') is not aligned and therefore may intersect with the edge. The neighborhoods of the pixels in S_2 may intersect the edge even though they are correctly aligned.

We further partition S_1 into P_1 and P_2 and S_3 into P_3 and P_4 such that

$$\begin{aligned}
P_1 &\triangleq \{(v,u) \mid h(v) + 2\delta_\ell + C/2\delta_s \leq u\},\\
P_2 &\triangleq \{(v,u) \mid h(v) + (1 + C/2)\delta_s \leq u \leq h(v) + 2\delta_\ell + C/2\delta_s\},\\
P_3 &\triangleq \{(v,u) \mid h(v) - (1 + C/2)\delta_s \geq u \geq h(v) - 2\delta_\ell - C/2\delta_s\},\\
P_4 &\triangleq \{(v,u) \mid u \leq h(v) - 2\delta_\ell - C/2\delta_s\}.
\end{aligned}$$

Lemma 1. *Any neighborhood of pixel $(v, u) \in P_1$ will lie completely above the edge contour.*

Proof. Let $(t_1, t_2) \in P_1$. If $(u, v) \in I_{\theta,\delta_s,\delta_\ell}(t_1, t_2)$, then

$$t_2 - v < \delta_\ell. \tag{12}$$

On the other hand, for $t_1' \in [t_1 - \delta_\ell, t_1 + \delta_\ell]$ we have,

$$h(t_1') = h(t_1) + h'(t_1)(t_1' - t_1) + \frac{1}{2}h''(t_1'')(t_1' - t_1)^2,$$

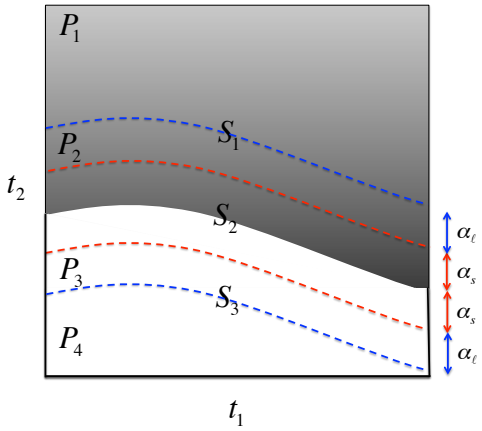


Figure 12: The regions P_1, P_2, P_3, P_4 with $\alpha_s = (1 + C/2)\delta_s$ and $\alpha_\ell = 2\delta_\ell - \delta_s$. Every neighborhood of $(t_1, t_2) \in P_1$ will lie completely above the edge contour. However, some of the neighborhoods of the pixels $(t_1, t_2) \in P_2$ may intersect the edge. A similar property holds for the regions P_3 and P_4.

where t_1'' is between t_1 and t_1'. Therefore,

$$h(t_1') - h(t_1) < \delta_\ell + \frac{1}{2} C\delta_s, \tag{13}$$

Comparing (12) and (13) completes the proof. □

In spite of P_1, some of the neighborhoods of the pixels $(v, u) \in P_2$ may intersect the edge. A similar property holds for P_3 and P_4, respectively. Figure 12 displays these regions. The following lemma, proved in [8], will be used in the proofs of the main theorems. For the sake of completeness, we sketch the proof here.

Lemma 2. *Let $Z_1, Z_2, \ldots, Z_r$ be iid $N(0,1)$ random variables. The χ^2_r random variable $\sum_{i=1}^r Z_i^2$ concentrates around its mean with high probability, i.e.,*

$$\mathbb{P}\left(\frac{1}{r}\sum_i Z_i^2 - 1 > t\right) \leq \mathrm{e}^{-\frac{r}{2}(t-\ln(1+t))}, \tag{14}$$

$$\mathbb{P}\left(\frac{1}{r}\sum_i Z_i^2 - 1 < -t\right) \leq \mathrm{e}^{-\frac{r}{2}(t+\ln(1-t))}. \tag{15}$$

Proof. Here we prove (14); the proof of (15) follows along very similar lines. From Markov's Inequality, we have

$$\begin{aligned}\mathbb{P}\left(\left(\frac{1}{r}\sum_{i=1}^{r} Z_i^2\right) - 1 > t\right) &\leq \mathrm{e}^{-\eta t-\eta}\mathbb{E}\left(\mathrm{e}^{\frac{\eta}{r}\sum_{i=1}^{r} Z_i^2}\right) \\ &= \mathrm{e}^{-\eta t-\eta}\left(\mathbb{E}\left(\mathrm{e}^{\frac{\eta Z_1^2}{r}}\right)\right)^r = \frac{\mathrm{e}^{-\eta t-\eta}}{\left(1-\frac{2\eta}{r}\right)^{\frac{r}{2}}}. \end{aligned} \tag{16}$$

The last inequality follows from Lemma 3 in [8]. The upper bound proved above holds for any $\eta < \frac{r}{2}$. To obtain the lowest upper bound we minimize $\frac{\mathrm{e}^{-\eta t-\eta}}{\left(1-\frac{2\eta}{r}\right)^{\frac{r}{2}}}$ over η. The optimal value is $\eta^\star = \arg\min_\eta \frac{\mathrm{e}^{-\eta t-\eta}}{\left(1-\frac{2\eta}{r}\right)^{\frac{r}{2}}} = \frac{rt}{2(t+1)}$. Plugging $\eta^\star$ into (16) proves the lemma. □

5.2. Proof of Theorem 2

In the set up above, we consider three different regions for the point (t_1, t_2). As we will see in the proof, the risk of all pixels in each region has the same upper bound. We calculate these upper bounds and then combine them to obtain a master upper bound for the risk of OANLM.

Case I: $(t_1, t_2) \in S_1$. We know that if the anisotropic neighborhood of (t_1, t_2), $I_{\theta,\delta_s,\delta_\ell}$ is aligned with the edge contour, i.e., $\tan(\theta) = h'(t_1)$, then it does not intersect the edge contour. To calculate the OANLM estimate we first calculate the weights. Define

$$\mathbf{z}_{t_1,t_2}^{\theta,\delta_s,\delta_\ell}(j_1, j_2) = \frac{n_s n_\ell}{\delta_s \delta_\ell}\int_{(s_1,s_2)\in I_{\theta,\delta_s,\delta_\ell}^{j_1,j_2}} dW(s_1, s_2), \tag{17}$$

and

$$\begin{aligned} Z(t_1, t_2) &= \frac{n_s n_\ell}{\delta_s \delta_\ell}\int\int_{(s_1,s_2)\in I_{\theta,\frac{\delta_s}{n_s},\frac{\delta_\ell}{n_\ell}}(t_1,t_2)} dW(s_1, s_2), \\ F(t_1, t_2) &= \frac{n_s n_\ell}{\delta_s \delta_\ell}\int\int_{(s_1,s_2)\in I_{\theta,\frac{\delta_s}{n_s},\frac{\delta_\ell}{n_\ell}}(t_1,t_2)} f(s_1, s_2)ds_1 ds_2. \end{aligned}$$

For notational simplicity we will use $w(s_1, s_2)$ and $\mathbf{z}_{t_1,t_2}(j_1, j_2)$ instead of $w^{\theta,\delta_s,\delta_\ell}_{t_1,t_2}(s_1, s_2)$ and $\mathbf{z}^{\theta,\delta_s,\delta_\ell}_{t_1,t_2}(j_1, j_2)$, respectively. Define

$$\begin{aligned} A_1 &\triangleq \{(u,v) \in P_1 \mid w(u,v) = 1\}, \\ A_2 &\triangleq \{(u,v) \in P_4 \mid w(u,v) = 0\}. \end{aligned}$$

Here, let $\lambda(\cdot)$ denote the Lebesgue measure of a set over $\mathbb{R}^2$.

Lemma 3. *Let* $\delta_\ell = 2\sigma^{2/3}|\log(\sigma)|^{2/3}$ *and* $\delta_s = 4\sigma^{4/3}|\log(\sigma)|^{4/3}$, $n_s = 2|\log(\sigma)|^{2/3}$ $n_\ell = |\log(\sigma)|^{4/3}$, *and* $\tau_\sigma = \frac{2}{\sqrt{|\log\sigma|}}$. *Then,*

$$\begin{aligned} \mathbb{P}(\lambda(P_1) - \lambda(A_1) > \epsilon) &= O\left(\frac{\sigma^8}{\epsilon}\right), \\ \mathbb{P}(\lambda(P_4) - \lambda(A_2) > \epsilon) &= O\left(\frac{\sigma^8}{\epsilon}\right). \end{aligned}$$

Proof. Here we prove the first inequality. The proof of the second inequality follows a similar route. Since $\mathbf{z}_{t_1,t_2}(j_1, j_2)$ is the integral of the Wiener sheet, we have $\mathbf{z}_{t_1,t_2}(j_1, j_2) \sim N(0, \frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell})$. Combining this with Lemma 2 we conclude that

$$\mathbb{P}\left(d^2_{\theta,\delta_s,\delta_\ell} - 2\frac{n_s n_\ell \sigma^2}{\delta_s \delta_\ell} \le t\right) \le \mathrm{e}^{-\frac{n_s n_\ell t^2}{4}} = O(\sigma^8).$$

Therefore,

$$\begin{aligned} \mathbb{E}(\lambda(A)) &= \mathbb{E}\int_{(u,v)\in P_1} I((u,v) \in A) = \int_{(u,v)\in P_1} \mathbb{P}((u,v) \in A) \\ &= \lambda(P_1) - O(\sigma^8). \end{aligned} \tag{18}$$

An upper bound for $\mathbb{P}(\lambda(P_1) - \lambda(A) > \epsilon)$ using the Markov inequality yields the result

$$\mathbb{P}(\lambda(P_1) - \lambda(A) > \epsilon) \le \frac{\mathbb{E}(\lambda(P_1) - \lambda(A))}{\epsilon} = O\left(\frac{\sigma^8}{\epsilon}\right).$$

□

Define the event $\mathcal{E}$ as

$$\mathcal{E} = \{\lambda(P_1) - \lambda(A_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(A_2) < \sigma^2\}.$$

For notational simplicity we also define the following notations:

$$\begin{aligned} wXdudv &\triangleq w(u,v)X(u,v)dudv, \\ Xdudv &\triangleq X(u,v)dudv, \\ wdudv &\triangleq w(u,v)dudv, \\ Zdudv &\triangleq Z(u,v)dudv, \\ wZdudv &\triangleq w(u,v)Z(u,v)dudv. \end{aligned}$$

The risk of the OANLM estimator at $(t_1, t_2) \in S_1$ is then given by

$$\mathbb{E}\left(\frac{\int_S wXdudv}{\int_S wdudv}\right)^2.$$

Define $U = \left(\frac{\int_S wXdudv}{\int_S wdudv}\right)^2$. We have

$$\mathbb{E}(U) = \mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) + \mathbb{E}(U \mid \mathcal{E}^c)\mathbb{P}(\mathcal{E}^c) \leq \mathbb{E}(U \mid \mathcal{E})P(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c). \quad (19)$$

Lemma 3 proves that $\mathbb{P}(\mathcal{E}^c) = O(\sigma^6)$. We have

$$\begin{aligned} &\left|\int_{P_1} wXdudv - \int_{P_1} Xdudv\right| \\ &= \left|\int_{P_1} wXdudv - \int_{A_1} wXdudv + \int_{A_1} Xdudv - \int_{P_1} Xdudv\right| \\ &= \left|\int_{P_1} wXdudv - \int_{A_1} wXdudv\right| + \left|\int_{A_1} Xdudv - \int_{P_1} Xdudv\right| \\ &= O(\lambda(P_1 \backslash A_1)). \end{aligned} \quad (20)$$

For the last inequality we have assume that the estimate is bounded over the P_1 region. This assumption can be also justified by calculating the probability that this condition does not hold and showing that the probability is negligible. Using arguments similar to (20) it is straightforward to prove that

$$\begin{aligned}
\left|\int_{P_4} wX dudv - \int_{P_4} X dudv\right| &= O(\lambda(P_4 \backslash A_2)). \\
\left|\int_{P_1} w dudv - \int_{P_1} dudv\right| &= O(\lambda(P_1 \backslash A_1)). \\
\left|\int_{P_4} w dudv - \int_{P_4} dudv\right| &= O(\lambda(P_4 \backslash A_2)). \qquad (21)
\end{aligned}$$

Define $P_{14} = P_1 \cup P_4$ and $B = S \backslash P_{14}$. We now calculate an upper bound for the first term of (19)

$$\begin{aligned}
&\mathbb{E}(U \mid \mathcal{E})\mathbb{P}(\mathcal{E}) \\
&= \mathbb{E}\left(\left(\frac{\int_{P_{14}} wX dudv + \int_B wX dudv}{\int_{P_{14}} w dudv + \int_B w dudv}\right)^2 \;\middle|\; \mathcal{E}\right)\mathbb{P}(\mathcal{E}) \\
&= \mathbb{E}\left(\left(\frac{\int_{P_1} X dudv + \int_B wX dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2 \;\middle|\; \mathcal{E}\right)\mathbb{P}(\mathcal{E}) + O(\sigma^2) \\
&\leq \mathbb{E}\left(\frac{\int_{P_1} X dudv + \int_B wX dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2 + O(\sigma^2) \\
&\leq \mathbb{E}\left(\frac{\int_B wF dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2 + \mathbb{E}\left(\frac{\int_{P_1} Z dudv + \int_B wZ dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2 \\
&\quad + 2\sqrt{\mathbb{E}\left(\frac{\int_{P_1} F dudv + \int_B wF dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2}\sqrt{\mathbb{E}\left(\frac{\int_{P_1} Z dudv + \int_B wZ dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2} \\
&\quad + O(\sigma^2). \qquad (22)
\end{aligned}$$

Lemma 4. *Let $w(u,v)$ be the weights of OANLM with $\delta_\ell, \delta_s, n_s, n_\ell$, and τ_σ as specified in Lemma 3. Then*

$$\mathbb{E}\left(\frac{\int_B wF dudv}{\int_{P_1} dudv + \int_B w dudv}\right)^2 \leq O(\sigma^{4/3}|\log(\sigma)|^{4/3}).$$

Proof. We have

$$\begin{aligned}\mathbb{E}\left(\frac{\int_B wF du dv}{\int_{P_1} du dv + \int_B w du dv}\right)^2 &\leq \mathbb{E}\left(\frac{\int_B du dv}{\int_{P_1} du dv}\right)^2 \\ = \quad \left(\frac{\lambda(B)}{\lambda(P_1)}\right)^2 &= O(\sigma^{4/3}|\log(\sigma)|^{4/3}).\end{aligned}$$

To obtain the inequality we maximize the numerator and minimize the denomerator independently. □

Lemma 5. *Let $w(u,v)$ be the weights of OANLM with $\delta_\ell, \delta_s, n_s, n_\ell$, and τ_σ as specified in Lemma 3. Then*

$$\mathbb{E}\left(\frac{\int_{P_1} Z du dv + \int_B wZ du dv}{\int_{P_1} du dv + \int_B w du dv}\right)^2 \leq O(\sigma^{4/3}|\log(\sigma)|^{4/3}).$$

Proof. Since $\int_B w du dv \geq 0$, and we are interested in the upper bound of the risk, we can remove it from the denominator to obtain

$$\begin{aligned}\mathbb{E}\left(\frac{\int_{P_1} Z du dv + \int_B wZ du dv}{\int_{P_1} du dv + \int_B w du dv}\right)^2 &\leq \mathbb{E}\left(\frac{\int_{P_1} Z du dv + \int_B wZ du dv}{\int_{P_1} du dv}\right)^2 \\ \leq \quad \underbrace{\mathbb{E}\left(\frac{\int_{P_1} Z du dv}{\int_{P_1} du dv}\right)^2}_{V_1} &+ \underbrace{\mathbb{E}\left(\frac{\int_B wZ du dv}{\int_{P_1} du dv}\right)^2}_{V_2} \\ + \quad \underbrace{2\sqrt{\mathbb{E}\left(\frac{\int_{P_1} Z du dv}{\int_{P_1} du dv}\right)^2}\sqrt{\mathbb{E}\left(\frac{\int_B wZ du dv}{\left(\int_{P_1} du dv\right)^2}\right)^2}}_{V_3}&.\end{aligned}$$

Now we obtain upper bounds for V_1,V_2, and V_3 separately. First, we have

$$\begin{aligned} V_1 &= \mathbb{E}\left(\frac{\int_{P_1} Z du dv}{\int_{P_1} du dv}\right)^2 = \mathbb{E}\left(\frac{\int_{P_1}\int_{P_1} Z(u,v)Z(u',v')du dv du' dv'}{\left(\int_{P_1} du dv\right)^2}\right) \\ &= \left(\frac{\int_{P_1}\int_{I_{\theta,\frac{2\delta_s}{n_s},\frac{2\delta_\ell}{n_\ell}}(u,v)} \mathbb{E}(Z(u,v)Z(u',v'))du dv du' dv'}{\left(\int_{P_1} du dv\right)^2}\right) \\ &\leq \left(\frac{\int_{P_1}\int_{I_{\theta,\frac{2\delta_s}{n_s},\frac{2\delta_\ell}{n_\ell}}(u,v)} \frac{n_s n_\ell \sigma^2}{\delta_s\delta_\ell} du dv du' dv'}{\left(\int_{P_1} du dv\right)^2}\right) = O\left(\frac{\delta_s\delta_\ell}{n_s n_\ell}\right). \end{aligned} \tag{23}$$

To obtain an upper bound for V_2, we first note that

$$\begin{aligned} &\mathbb{E}(w(u,v)Z(u,v)w(u',v')Z(u',v')) \\ &\quad\leq \sqrt{\mathbb{E}(w(u,v)Z(u,v))^2}\sqrt{\mathbb{E}(w(u',v')Z(u',v'))^2} = \frac{n_s n_\ell \sigma^2}{\delta_s\delta_\ell}, \end{aligned} \tag{24}$$

and hence,

$$\begin{aligned} V_2 &= \mathbb{E}\left(\frac{\int_B wZ du dv}{\int_{P_1} du dv}\right)^2 \\ &= \mathbb{E}\left(\frac{\int_B\int_B w(u,v)Z(u,v)w(u',v')Z(u',v')du dv du' dv'}{(\int_{P_1} du dv)^2}\right) \\ &\leq \left(\frac{\int_B\int_B \frac{n_s n_\ell\sigma^2}{\delta_s\delta_\ell} du dv du' dv'}{(\int_{P_1} du dv)^2}\right) = O(\delta_\ell^2). \end{aligned} \tag{25}$$

Using (23) and (25) it is straightforward to obtain an upper bound for V_3. Combining the upper bounds for V_1, V_2, and V_3 completes the proof. □

Combining (22) with Lemmas 4 and 5 establishes the following upper bound for Case I, where $(t_1, t_2) \in S_1$:

$$\mathbb{E}\left(f(t_1,t_2) - \frac{\int_S w(u,v)X(u,v)du dv}{\int_S w(u,v)du dv}\right)^2 = O(\sigma^{4/3}|\log(\sigma)|^{4/3}). \tag{26}$$

Case II: $(t_1, t_2) \in S_2$. In this case we assume that the risk is bounded by 1, since the function f is bounded between 0 and 1. If the estimate is out of this range, then we will map it to either 0 or 1.

Case III: $(t_1, t_2) \in S_3$. This case is exactly the same as Case I , and hence we skip the proof.

Finally, combining our results for Cases I, II, and III, we can calculate an upper bound for the risk of OANLM as

$$\begin{aligned} R(f, \hat{f}^o) &= \mathbb{E}(\|f - \hat{f}^o\|^2) = \int_{S_1 \cup S_3} (f - \hat{f}^o)^2 dt_1 dt_2 + \int_{S_2} (f - \hat{f}^o)^2 dt_1 dt_2 \\ &= \lambda(S_1 \cup S_3) O(\sigma^{4/3} |\log(\sigma)|^{4/3}) + \lambda(S_2) O(1) = O(\sigma^{4/3} |\log(\sigma)|^{4/3}). \end{aligned}$$

So ends the proof of Theorem 2.

5.3. Proof of Corollary 1

The proof of this corollary follows exactly the same route as that of Theorem 2. We merely redefine the regions S_k and P_k for $k \in \{1, 2, 3\}$. The new definition of the S_k regions for the NLM algorithm is given by

$$\begin{aligned} S_1^n &= \{(t_1, t_2) \mid t_2 > h(t_1) + 2\delta\}, \\ S_3^n &= \{(t_1, t_2) \mid t_2 < h(t_1) - 2\delta\}, \end{aligned}$$

and $S_2^n = S \backslash S_1^n \cup S_3^n$. Since the neighborhoods in NLM are isotropic, we require a single parameter to describe the neighborhood size, $\delta = \delta_s = \delta_\ell$. The main assumption is that $\frac{C}{2}\delta^2 < \sigma$. This is clear since δ is assumed to go to zero as $\sigma \to 0$. Therefore in the isotropic case, the neighborhood of the pixels in S_1 and S_3 does not intersect with the edge contour. Since the neighborhoods no longer have different anisotropic lengths, the size of the intersection of the neighborhoods and the edge contour is identical in all directions and there is no need to define the regions, $P_1, \ldots, P_4$. Equivalently, $P_1 = S_1$, $P_4 = S_4$ and $P_2 = P_3 = \emptyset$. With these definitions the proof of this corollary is exactly the same as above.

5.4. Proof of Theorem 3

The proof of this theorem is very similar to the proof of Theorem 2. The only difference is in the definitions of S_k and P_k for $k \in \{1, 2, 3\}$. Since

there is a mismatch between the orientations of the neighborhood and the edge contour, the neighborhood of S_1 may intersect the edge. In order to fix this, we define the new regions called S_i^β and P_i^β. If the error in θ is upper bounded by $c_\beta \sigma^\beta$, then define

$$\begin{aligned} S_1^\beta &= \{(t_1,t_2) \mid t_2 > h(t_1) + c_\beta\sigma^\beta\delta_\ell + \delta_s + C/2\delta_\ell^2\}, \\ S_3^\beta &= \{(t_1,t_2) \mid t_2 < h(t_1) - c_\beta\sigma^\beta\delta_\ell - \delta_s - C/2\delta_\ell^2\}, \end{aligned}$$

and $S_2^\beta = S \backslash S_1^\beta \cup S_3^\beta$. Furthermore, define $P_1^\beta = P_1$, $P_4^\beta = P_4$, $P_2^\beta = S_1^\beta \backslash P_1$, and $P_3^\beta = S_3^\beta \backslash P_4$. Using these new partitions, the proof follows exactly the same route as the proof of Theorem 2.

5.5. Proof of Theorem 4

Since the proof is mostly similar to that of Theorem 2, we shall focus on the major differences. The first difference is that we consider the regions P_1–P_4 instead of S_1–S_3. Previously the regions P_1 and P_2 were treated jointly under the region S_1. Instead, we shall now consider P_1 and P_2 separately, and their differences shall become progressively apparent.

Case I: $(t_1,t_2) \in P_1$. We start with the calculation of the weights $w(u,v)$ for $(u,v) \in P_1 \cup P_4$. Define

$$\begin{aligned} R_1 &\triangleq \{(u,v) \in P_1 \mid w^D(u,v) = 1\}, \\ R_2 &\triangleq \{(u,v) \in P_4 \mid w^D(u,v) = 0\}. \end{aligned}$$

Lemma 6. *Let $\delta_s, \delta_\ell, n_s, n_\ell, t$ be as defined in Lemma 3 and let $q = \pi\sigma^{-2/3}$ in the DANLM algorithm. Then,*

$$\begin{aligned} \mathbb{P}(\lambda(P_1) - \lambda(R_1) > \epsilon) &\leq \frac{\pi\sigma^{22/3}}{\epsilon}, \\ \mathbb{P}(\lambda(P_4) - \lambda(R_2) > \epsilon) &\leq \frac{\pi\sigma^{22/3}}{\epsilon}. \end{aligned}$$

The proof is a simple application of Lemma 3 and the union bound. Note that since $(t_1,t_2) \in P_1$, all of its directional neighborhoods are located above the edge and do not intersect with the edge contour.

Define the event $\mathcal{F}$ as

$$\mathcal{F} = \{\lambda(P_1) - \lambda(R_1) < \sigma^2\} \cap \{\lambda(P_4) - \lambda(A_4) < \sigma^2\}.$$

Lemma 7 confirms that $\mathbb{P}(\mathcal{F}^c) = O(\sigma^{16/3})$. The rest of the proof of Case I is exactly the same as the proof of Case I in Theorem 2, and therefore we do not repeat it here.

Case II: $(t_1, t_2) \in P_2$. In this case we again start with defining the following two sets:

$$\begin{aligned} L_1 &\triangleq \{(u,v) \in P_1 \mid w^D_{u,v} = 1\}, \\ L_2 &\triangleq \{(u,v) \in P_4 \mid w^D_{u,v} = 0\}. \end{aligned}$$

Lemma 7. *Let $\delta_s, \delta_\ell, n_s, n_\ell, \tau_\sigma$ be as defined in Lemma 3, and let $q = \pi\sigma^{-2/3}$. Then,*

$$\mathbb{P}(\lambda(P_1) - \lambda(L_1) > \epsilon) \leq \frac{\pi\sigma^{22/3}}{\epsilon}.$$

Proof. If we prove that for some $\theta \in \Theta$, $I_{\theta,\delta_s,\delta_\ell}(t_1, t_2)$ does not intersect the edge, then the proof of the lemma is immediate. Suppose that θ^* is such that $\tan(\theta^*) = h'(t_1)$, and consider $\hat{\theta} = \arg\min_{\theta\in\Theta} |\theta - \theta^*|$. To ensure that the neighborhood $I_{\theta,\delta_s,\delta_\ell}(t_1, t_2)$ does not intersect with the edge, we need to have

$$\begin{aligned} t_2 + \tan(\hat{\theta})\,(t' - t_1) - \delta_s \;>\; h(t_1) &+ h'(t_1)\,(t' - t_1) \\ &+ \frac{1}{2}h''(t'')\left(t' - \frac{i}{n}\right)^2 \end{aligned} \tag{27}$$

for all $t' \in [t_1, t_1 + \delta_\ell)$. Also, $t'' \in [t_1, t']$. Clearly, $\tan(\hat{\theta}) = \tan(\theta^*) + \frac{1}{1+\tan^2(\theta')}\Delta\theta$, where $\Delta\theta = \hat{\theta} - \theta^*$. Furthermore, $|h''(t'')| \leq C_2$ and $\Delta\theta \leq \frac{\pi}{2}\delta_\ell$. Therefore, (27) will be satisfied if

$$t_2 > h\,(t_1) + \frac{\pi\delta_\ell}{2}\delta_\ell + \delta_s + C_2\delta_\ell^2.$$

This constraint holds for all pixels in the region P_2. Therefore, at least one of the neighborhoods is completely in the region above h, which in turns proves the lemma. □

Lemma 8. *Let $\delta_s, \delta_\ell, n_s, n_\ell, \tau_\sigma$ be as defined in Lemma 3, and let $q = \pi\sigma^{-2/3}$. Then,*

$$\mathbb{P}(\lambda(P_4) - \lambda(L_2) > \epsilon) \leq \frac{\pi\sigma^{10/3}}{\epsilon}.$$

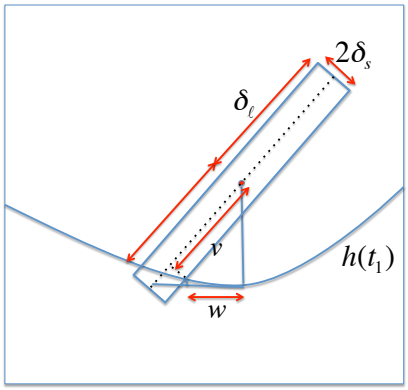


Figure 13: Explanation of the neighborhood rotation in Lemma 8.

Proof. We first prove that at least half of the area of $I_{\theta,\delta_s,\delta_\ell}(u,v)$ is located below the edge. For notational simplicity we assume that $h'(t_1)=0$. Consider Figure 13, in which the neighborhood has an arbitrary orientation. Let w, v be as defined in this figure. We then have that

$$v = \frac{d}{\cos\theta} - \frac{Cw^2}{\cos\theta} - \delta_s \tan\theta = \frac{d - Cw^2 - \delta_s \sin\theta}{\cos\theta} \geq \frac{d - Cw^2 - \delta_s}{\cos\theta}, \quad (28)$$

where $d = t_2 - h(t_1)$. Suppose d is such that, at the correct angle $d - Cw^2 - \delta_s > 0$, i.e., $d > (C+1)\sigma^{4/3}$. According to (28), $v > 0$, and hence more than half of the area of the neighborhood is in the white region. Therefore, the number of pixels in this region is $4\log^2\sigma \pm o(\log^2\sigma)$. □

Case III: $(t_1, t_2) \in S_2$. The proof of this case is identical to that of Case II in the proof of Theorem 2.

Case IV: $(t_1, t_2) \in P_3 \cup P_4$. The proof of this case is similar to that for the regions P_1 and P_2 in Cases I and II and therefore is skipped here.

Finally, combining the upper bounds obtained in Cases I, II, III, and IV completes the proof of the theorem.

6. Related work on anisotropic denoising

Anisotropy is a fundamental feature of real-world images. As a result, anisotropic image processing tools can be traced back at least as far as

the 1970s [29]. Here, we briefly compare and contrast some of the relevant anisotrpic denoising schemes with ANLM.

Anisotropic filtering methods use a space-varying linear convolution filters to reduce blur along image edges. Toward this goal, [29] considers several different neighborhoods around a pixel and selects the most "homogeneous" neighborhood to smooth over. A more advanced version of this idea can be found in [30]. There are major differences between such algorithms and NLM/ANLM, in particular, the estimators are local and do not exploit global similarities.

Anisotropic diffusion methods smooth a noisy image with a Gaussian kernel. As the standard deviation of the kernel increases, the smoothing process introduces larger bias to the edges. In [31, 32] the authors proved that the set of images derived by this approach can be viewed as the solution to the heat diffusion equation. Perona and Malik [33] noted the isotropy in the heat equation and introduced anisotropy. Their anisotropic diffusion starts with the heat equation but at every iteration exploits the gradient information of the previous estimate to increase the conductance along the edges and decrease it across the edges. Efforts to theoretically analyze the risk of this algorithm have left many open questions remaining [14]. It is worth noting that the idea of applying an image denoising algorithm iteratively and guiding it at every iteration based on previous estimates goes back to Tukey's twicing [34].

Anisotropic transformations enable simple threshold based denoising algorithms. While the standard separable wavelet transform cannot exploit the smoothness of the edge contour, a menagerie of anisotropic transforms have been developed, including ridgelets [15], curvelets [16], wedgelets [9], platelets [17, 35], shearlets [18], contourlets [19], bandelets [20], and directionlets [21]. As mentioned in the Introduction, among the above algorithms only wedgelets can obtain the optimal minimax risk for the Horizon class; however wedgelets are not suited to denoising textures. One promising avenue combing wavelets (for texture denoising) and wedgelets (for edge denoising) could follow the path of the image coder in [36].

Alternative *anisotropic NLM algorithms* have been proposed to address the inefficiency of using a fixed neighborhood. In [37, 38] the authors adapt the neighborhood size to the local image content but do not provide any optimality analysis. In [38] the authors consider different isotropic NLM

neighborhood sizes depending on how smoothly the image content varies. In [37] the authors do not change the neighborhood size (which is critical to acheive optimality) but rather use image gradient information to increase the weights of the NLM along edges and decrease them across edges. This method is equivalent to modifying the threshold parameter t_n to force NLM to assign higher weights to edge-like neighborhoods. Unfortunately, this technique does not reduce the bias that renders NLM sub-optimal.

Finally, *data-driven optimality criteria* have been considered in [39–41], where the authors derive lower bounds for the performance of denoising algorithms. However, the analyses provided in these papers are not fully rigorous and do not cover the performance of NLM for images with sharp edges.

7. Discussion and future directions

We have introduced and analyzed a framework for anisotropic nonlocal means (ANLM) denoising. Similar to NLM, ANLM exploits nonlocal information in estimating the pixel values. However, unlike NLM, ANLM uses anisotropic, oriented neighborhoods that can exploit the smoothness of edge contours. This enables ANLM to outperform NLM both theoretically and empirically. In fact, the performance of ANLM is withtin a logarithmic factor of optimal as measured by the minimax rate on the Horizon class of images.

Numerous questions remain open for future research. From the theoretical perspective, the risk analysis of GANLM, the application to noise models beyond Gaussian, and the extension to three dimensions and beyond (for seismic, medical, and other data) pose interesting research challenges. From the practical perspective, the question of how to best tune ANLM to match the nuanced edges and textures of real-world images remains open, since we have considered only brutal binary images here. Finally, while NLM is no longer the state-of-the-art denoising algorithm, it is a key building block in several top-performing algorithms. It would be interesting to see whether anisotropy pays off as handily for those algorithms as it does for NLM.

8. Acknowledgements

This work was supported by the grants NSF CCF-0431150, CCF-0926127, and CCF-1117939; DARPA/ONR N66001-11-C-4092 and N66001-11-1-4090;

ONR N00014-08-1-1112, N00014-10-1-0989, and N00014-11-1-0714; AFOSR FA9550-09-1-0432; ARO MURI W911NF-07-1-0185 and W911NF-09-1-0383; and the TI Leadership University Program.

Appendix A. Minimax risk of the mean filter

In this appendix, we obtain the decay rate of the risk of the mean filter. Our proof is similar to the proof in [14]. Nevertheless, we repeat the proof here for the sake of completeness and since our continuous framework is slightly different from the discrete framework in [14]. See [8] for the generalization of this result.

The classical mean filter estimates the image via

$$\hat{f}^{\mathrm{MF}}(t_1,t_2)=\frac{1}{\Delta^2}\int_{\tau_1=-\frac{\Delta}{2}}^{\frac{\Delta}{2}}\int_{\tau_2=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} dY(t_1-\tau_1,t_2-\tau_2),$$

where Δ specifies the the size of the window on which averaging takes place.

Theorem 5 ([14])**.** *If $\hat{f}^{\mathrm{MF}}$ is the estimate of the mean filter, then*

$$\inf_{\Delta}\sup_{f\in\mathcal{H}^{\alpha}(C)} R(f,\hat{f}^{MF})=\Theta(\sigma^{2/3}).$$

Proof. We first derive a lower bound for $R(f,\hat{f}^{\mathrm{MF}})$. Consider a function $f_h(t_1,t_2)$ with $h(t_1)=1/2$ (recall Figure 2) and define

$$Q_{\Delta}=\{(t_1,t_2)\mid 1/2-\Delta<t_2<1/2+\Delta\}.$$

It is straightforward to confirm that if $(t_1,t_2)\in Q_{\Delta/4}$, then $|f_h(t_1,t_2)-\mathbb{E}\hat{f}^{\mathrm{MF}}(t_1,t_2)|>\frac{1}{4}$. Therefore, if $\mathrm{Bias}(\hat{f}^{\mathrm{MF}})$ denotes the bias of the mean filter estimator, then we have

$$\begin{aligned}\mathrm{Bias}^2(\hat{f}^{\mathrm{MF}}) &= \int_S |f_h(t_1,t_2)-\mathbb{E}\hat{f}^{\mathrm{MF}}(t_1,t_2)|^2dt_1dt_2\\ &\geq \int_{Q_{\Delta/4}} |f_h(t_1,t_2)-\mathbb{E}\hat{f}^{\mathrm{MF}}(t_1,t_2)|^2dt_1dt_2\geq\frac{1}{32}\Delta. \quad \text{(A.1)}\end{aligned}$$

Now consider a point $(t_1, t_2) \in S \backslash Q_\Delta$. We have

$$\begin{aligned}
&\mathbb{E}(\hat{f}^{\mathrm{MF}}(t_1,t_1) - \mathbb{E}\hat{f}^{\mathrm{MF}}(t_1,t_2))^2 \\
&= \int_{\tau_1=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau_2=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau_1'=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \int_{\tau_2'=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \frac{\mathbb{E}(dW(t_1-\tau_1, t_2-\tau_2) dW(t_1-\tau_1', t_2-\tau_2'))}{\Delta^4} \\
&= \frac{\sigma^2}{\Delta^2}
\end{aligned}$$

Therefore, if $\mathrm{Var}(\hat{f}^{\mathrm{MF}})$ is the variance of the mean filter estimator, then we have

$$\mathrm{Var}(\hat{f}^{\mathrm{MF}}) \geq \int_{S \backslash Q_{4\Delta}} \mathbb{E}(\hat{f}^{\mathrm{MF}}(t_1,t_1) - \mathbb{E}\hat{f}^{\mathrm{MF}}(t_1,t_2))^2 = \frac{\sigma^2}{\Delta^2}(1-4\Delta). \quad \text{(A.2)}$$

By combining (A.1) and (A.2) we obtain the lower bound of $\frac{\sigma^2}{\Delta^2} + \frac{1}{32}\Delta - 4\frac{\sigma^2}{\Delta}$ for the risk of the mean filter estimator. If we minimize this lower bound over Δ then we obtain the lower bound of $\Theta(\sigma^{2/3})$ for the risk.

Now, we derive an upper bound for the risk. Define the region

$$R_\Delta = \{(t_1, t_2) \mid h(t_1) - \Delta \leq t_2 \leq h(t_1) + \Delta\}.$$

It is straightforward to confirm that, if $(t_1, t_2) \in S \backslash R_\Delta$, then the Δ-neighborhood of this point does not intersect the edge contour. Therefore, the bias of the estimator over this region is zero and the variance is $\frac{\sigma^2}{\Delta^2}$. If we bound the risk of the points in the R_Δ region by 1, we obtain the following upper bound for the risk,

$$R(f, \hat{f}^{\mathrm{MF}}) \leq \frac{\sigma^2}{\Delta^2} + 2\Delta.$$

Again by optimizing the upper bound over Δ we obtain the upper bound of $\Theta(\sigma^{2/3})$. This completes the proof. □

Appendix B. Minimax risk of NLM

Corollary 1 states the following upper bound for the risk of NLM:

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^N) = O(\sigma |\log \sigma|).$$

In this section we prove that the risk of NLM is lower bounded by $\Omega(\sigma)$. The proof is similar to the proof of Theorem 5 in [8]; we consider the function $f_h(t_1, t_2)$ for $h(t_1) = 1/2$ and prove that the bias of the NLM on $(t_1, t_2) \in Q_{\frac{\delta}{n}}$ is $\Theta(1)$ for any choice of the threshold parameter. However, in the continuous setting considered here, the steps are more challenging. Following [8] we consider the semi-oracle NLM algorithm (SNLM). The semi-oracle δ-distance is defined as

$$\begin{aligned}&\tilde{d}^2_\delta(dY(t_1, t_2), dY(s_1, s_2))\\ &\quad = \frac{1}{n^2 - 1}\left(\|\mathbf{x}^\delta_{t_1,t_2} - \mathbf{y}^\delta_{s_1,s_2}\|_2^2 - |\mathbf{x}^\delta_{t_1,t_2}(0,0) - \mathbf{y}^\delta_{s_1,s_2}(0,0)|^2\right),\end{aligned}$$

where

$$\mathbf{x}^\delta_{t_1,t_2}(j_1, j_2) = \int_{(s_1,s_2)\in I^{j_1,j_2}_{\frac{\delta}{n}}} f(s_1, s_2)ds_1ds_2.$$

SNLM then estimates the weights according to

$$w^S_{t_1,t_2}(s_1, s_2) = \begin{cases} 1 & \text{if } \tilde{d}^2_\delta(dY(t_1, t_2), dY(s_1, s_2)) \le \frac{n^2\sigma^2}{\delta^2} + \tau_\sigma, \\ 0 & \text{otherwise.}\end{cases}$$

It is clear that the distance estimates of the SNLM algorithm are more accurate than those of NLM algorithm. Hence it outperforms NLM, and a lower bound that holds for SNLM will hold for NLM as well.

We make the following mild assumptions on the parameters of SNLM.

A1: The window size $\delta \to 0$ as $\sigma \to 0$.

A2: $\frac{\delta^2}{n^2} = \Omega(\sigma^2)$. Otherwise $\mathbb{E}\tilde{d}^2_\delta(dY(t_1, t_2), dY(s_1, s_2)) \to \infty$.

A3: $n \to \infty$ as $\sigma \to 0$. This ensures that, if two points (t_1, t_2) and (s_1, s_2) have the same neighborhoods, then $w_{t_1,t_2}(u_1, u_2) = 1$ with high probability.

A4: If $\tilde{d}(f(t_1, t_2), f(s_1, s_2)) > 1/4$, then $\mathbb{P}(w^S_{t_1,t_2}(s_1, s_2) = 1) = o(\sigma^3)$.

Again, consider the function $f_h(t_1, t_2)$ for $h(t_1) = 1/2$ (recall Figure 2). Let $(t_1, t_2) \in Q_{\delta/n}$. For notational simplicity we use $w(s_1, s_2)$ instead of $w^S_{t_1,t_2}(s_1, s_2)$.

Lemma 9. *If $|s_1 - t_1| > \delta/2$ and $|s_1' - t_1| > \delta/2$, then*

$$\mathbb{P}(w^S_{t_1,t_2}(s_1, s_2) = 1) = \mathbb{P}(w^S_{t_1,t_2}(s_1', s_2) = 1)$$

for any t_1, t_2, s_1, s_2 and s_1'.

The proof is straightforward and hence skipped here.

Now consider a point $(t_1, t_2) \in G_{\delta/n}$.

Lemma 10. *If $|s - t_1| > \delta/2$, for $u < \delta/4$, then we have*

$$\mathbb{P}(w^S_{t_1,t_2}(s, t_2 - u) = 1) = \mathbb{P}(w^S_{t_1,t_2}(s, t_2 + u) = 1)$$

This lemma is a straightforward application of symmetry, and hence we skip the proof.

Lemma 11. *Suppose that δ and τ_σ satisfy A1–A4. Then we have*

$$\mathbb{P}\left(\int_{S\backslash Q_{\delta/2}} w(s_1, s_2) ds_1 ds_2 > \sigma^2\right) = o(\sigma).$$

Proof. Define

$$B = \{(s_1, s_2) \in S\backslash Q_{\delta/2} \mid w(s_1, s_2) = 1\}$$

and the event

$$\mathcal{G} = \{\lambda(B) \geq \sigma^2\}.$$

Using the Markov Inequality and Assumption A4, it is straightforward to show that

$$\mathbb{P}(\mathcal{G}) = \mathbb{P}\left(\int_{S\backslash Q_{\delta/2}} w(u, v) du dv > \sigma^2\right) = o(\sigma\lambda(S\backslash Q_{\delta/2})) = o(\sigma).$$

On the other hand, if $\lambda(B) < \sigma^2$, then

$$\int_{S\backslash Q_{\delta/2}} w(u, v) du dv = O(\sigma^2).$$

This completes the proof. □

Proposition 1. *Let $(t_1, t_2), (s_1, s_2) \in Q_{\delta/n}$. Then there exists $p_0 > 0$ independent of σ such that for every t_n, δ we have*

$$\mathbb{P}(w^n_{t_1,t_2}(u, v) = 1) > p_0.$$

Proof. We need to demonstrate that regions above and below the edge contour are included in the NLM algorithm with a constant non-zero probability p_0. First, we use the definition of the weights of the NLM algorithm to determine the probability that $w(u, v) = 1$

$$\begin{aligned}
\mathbb{P}(w^n_{t_1,t_2}(s_1, s_2) = 1) &= \mathbb{P}\left(\tilde{d}^2_\delta(dY(t_1, t_2), dY(s_1, s_2) < \frac{n^2\sigma^2}{\delta^2} + \tau_\sigma\right) \\
&= \mathbb{P}\left(\frac{1}{n^2-1}\sum\left(s^2_{\ell,p} - \frac{n^2\sigma^2}{\delta^2}\right) - \frac{1}{n^2-1}\sum s_{\ell,0} \leq -\frac{1}{n} + \tau_\sigma\right) \\
&\geq \mathbb{P}\left(\frac{1}{n^2-1}\sum\left(s^2_{\ell,p} - \frac{n^2\sigma^2}{\delta^2}\right) - \frac{1}{n^2-1}\sum s_{\ell,0} \leq -\frac{1}{n}\right),
\end{aligned}$$

where $s_{\ell,p} = \mathbf{z}^{0,\delta,\delta}_{s_1,s_2}(\ell, p)$. Using the Berry-Esseen Central Limit Theorem for independent non-identically distributed random variables [42], we can easily bound this probability away from zero. For more details, see Proposition 1 in [8]. □

We now consider the weights for the regions above and below the edge contour separately. Therefore, we define

$$\begin{aligned}
Q^1_\Delta &= \{(t_1, t_2) \mid 1/2 < t_2 < 1/2 + \Delta/4\}, \\
Q^2_\Delta &= \{(t_1, t_2) \mid 1/2 - \Delta/4 < t_2 < 1/2\}, \\
\Omega^1_{\Delta,\delta} &= \{(t_1, t_2) \mid 1/2 < t_2 < 1/2 + \Delta/4,\ 0 < t_1 < 2\delta\}, \\
\Omega^2_{\Delta,\delta} &= \{(t_1, t_2) \mid 1/2 - \Delta/4 < t_2 < 1/2,\ 0 < t_1 < 2\delta\}.
\end{aligned}$$

Let

$$p_{u,\sigma} = \mathbb{P}(w^n_{(t_1,t_2)}(u, v) = 1).$$

Note that according to Lemma 9 this probability does not depend on v.

Lemma 12. *If $w(u, v)$ denotes the weights used in the NLM algorithm, then*

we have

$$\begin{aligned}
\mathbb{P}\left(\left|\int_{Q^1_\Delta} w(u,v)dudv - \int_{Q^1_\Delta} p_{u,\sigma}du\right| > 2\epsilon + 2\sqrt{\frac{|\log(\sigma)|}{\delta}}\Delta\delta\right) &= O\left(\frac{\sigma^4}{\epsilon}\right), \\
\mathbb{P}\left(\left|\int_{Q^2_\Delta} w(u,v)dudv - \int_{Q^2_\Delta} p_{u,\sigma}du\right| > 2\epsilon + 2\sqrt{\frac{|\log(\sigma)|}{\delta}}\Delta\delta\right) &= O\left(\frac{\sigma^4}{\epsilon}\right).
\end{aligned}$$

Proof. The weights $w(u,v)$ over the pixelated neighborhoods can be rewritten in the following way

$$\int_{Q^1_\Delta} w(u,v)dudv = \int_{\Omega^1_\Delta} \sum_{k=1}^{\frac{1}{2\delta}} w(u+2k\delta, v)dudv.$$

Applying the Hoeffding inequality, it is straightforward to confirm that

$$\mathbb{P}\left(\left|\sum_{k=1}^{\frac{1}{2\delta}} w(u+2k\delta,v) - \frac{p_{u,\sigma}}{2\delta}\right| > t\right) \leq 2\mathrm{e}^{-2\delta t^2}.$$

Now, defining

$$\Gamma^1_\Delta = \left\{(u,v) \in \Omega^1_\Delta \;\middle|\; \left|\sum_{k=1}^{\frac{1}{2\delta}} w(u+2k\delta,v) - \frac{p_{u,\sigma}}{2\delta}\right| < t\right\},$$

we see that

$$\begin{aligned}
\mathbb{P}(\lambda(\Omega^1_\Delta) - \lambda(\Gamma^1_\Delta) > \epsilon) &= \mathbb{P}\left(\int_{(u,v)\in\Omega^1_\Delta} (1 - \mathbb{I}((u,v) \in \Gamma^1_\Delta)dudv > \epsilon\right) \\
&\leq \frac{2\mathrm{e}^{-2\delta t^2}}{\epsilon},
\end{aligned}$$

where the last step is due to Markov's Inequality. Define the event

$$\mathcal{F} = \{\lambda(\Omega^2_\Delta) - \lambda(\Gamma^2_\Delta) < \epsilon\}.$$

If $\lambda(\Omega^2_\Delta) - \lambda(\Gamma^2_\Delta) < \epsilon$, then we have that

$$\left|\int_{Q^1_\Delta} w(u,v)dudv - \int_{Q^1_\Delta} p(u,\sigma)dudv\right| \leq \left|\int_{Q^1_\Delta} w(u,v)dudv - \int_{\Gamma^1_\Delta} w(u,v)dudv\right|$$
$$+ \left|\int_{\Gamma^1_\Delta} w(u,v)dudv - \int_{\Gamma^1_\Delta} p(u,\sigma)dudv\right| + \left|\int_{\Gamma^1_\Delta} p(u,\sigma)dudv - \int_{Q^1_\Delta} p(u,\sigma)dudv\right|$$
$$\leq \quad 2\epsilon + 2t\Delta\delta.$$

Setting $t = \sqrt{\frac{|\log\sigma|}{\delta}}$ completes the proof. □

Theorem 6. *Suppose that δ, τ_σ and n satisfy Assumptions A1–A4. Then the risk of SNLM is*

$$\inf_{\delta_n, t_n} \sup_{f \in H^\alpha(C)} R(f, \hat{f}^S) = \Omega(\sigma).$$

Proof. Let $(t_1, t_2) \in \Omega^2_{\delta/n}$. We will calculate the bias of the NLM estimator at this point. We have

$$\mathbb{E}\left(f(t_1,t_2) - \frac{\int_S w(u,v)X(u,v)dudv}{\int_S w(u,v)dudv}\right)^2 \geq \left(\mathbb{E}\left(\frac{\int_S w(u,v)X(u,v)dudv}{\int_S w(u,v)dudv}\right)\right)^2.$$

which leads us to towards the lower bound

$$\begin{aligned}
\mathbb{E}\left(\frac{\int_S w(u,v)X(u,v)dudv}{\int_S w(u,v)dudv}\right) &\geq \mathbb{E}\left(\frac{\int_S w(u,v)X(u,v)dudv}{\int_S w(u,v)dudv} \mid \mathcal{F}\cap\mathcal{G}\right)\mathbb{P}(\mathcal{F}\cap\mathcal{G}) \\
&\geq \mathbb{E}\left(\frac{\int_{Q_{\delta/4}} w(u,v)X(u,v)dudv}{\int_{Q_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2} \mid \mathcal{F}\cap\mathcal{G}\right)\mathbb{P}(\mathcal{F}\cap\mathcal{G}) \\
&\geq \mathbb{E}\left(\frac{\int_{Q_{\delta/4}} w(u,v)X(u,v)dudv}{\int_{Q_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2}\right) - \mathbb{P}(\mathcal{F}^c\cup\mathcal{G}^c) \\
&= \mathbb{E}\left(\frac{\int_{Q_{\delta/4}} w(u,v)f(u,v)dudv}{\int_{Q_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2}\right) - \mathbb{P}(\mathcal{F}^c\cup\mathcal{G}^c) \\
&= \mathbb{E}\left(\frac{\int_{Q^1_{\delta/4}} w(u,v)dudv}{\int_{Q_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2}\right) - \mathbb{P}(\mathcal{F}^c\cup\mathcal{G}^c) \\
&= \mathbb{E}\left(\frac{\int_{Q^1_{\delta/4}} p(u,v)dudv + 2\epsilon + t}{\int_{Q_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2}\right) - \mathbb{P}(\mathcal{F}^c\cup\mathcal{G}^c) \\
&\geq \mathbb{E}\left(\frac{\int_{Q^1_{\delta/4}} p(u,v)dudv + 2\epsilon + t}{\sigma + \int_{2Q^1_{\delta/4}} p(u,v)dudv - 2\epsilon - t\delta^2}\right) - \mathbb{P}(\mathcal{F}^c\cup\mathcal{G}^c). \qquad \text{(B.1)}
\end{aligned}$$

The minimum of the last line is achieved when $\int_{Q^1_{\delta/4}} p(u,v)dudv$ is minimized. However, we have proved in Proposition 1 that $p(u) > p_0$ for $(u,v) \in Q^1_{\delta/n}$. Therefore, this minimizing integral is $\Omega(\sigma)$. When we substitute this optimum value in the lower bound (B.1), we see that the risk over this region is $\Theta(1)$. Therefore the bias of the NLM over the entire image is $\Omega(\frac{\delta}{n})$ or, equivalently, $\Omega(\sigma)$ according to Assumption A2. □

Appendix C. Proof of Theorem 1

Here we focus on the case of $\alpha = 2$ and use the standard technique of hypercube construction to establish the lower bound [12]. Let $\phi : [0,1] \to \mathbb{R}^+ \cup \{0\}$ be a two times differentiable function with $\|\frac{d^2\phi}{dt^2}\|_\infty = 1$, $\phi(0) = \phi(1) = 0$, $\frac{d\phi}{dt}\big|_0 = \frac{d\phi}{dt}\big|_1 = 0$, and $\frac{d^2\phi}{dt^2}\Big|_0 = \frac{d^2\phi}{dt^2}\Big|_1 = 0$. Define

$$\phi_{i,m} \triangleq Cm^{-2}\phi(mt - i) \qquad i = 0, 1, \ldots, m-1.$$

Set $f_0 \triangleq \mathbf{1}_{\{t_2<0.5\}}$, and define $\psi_{i,m} \triangleq \mathbf{1}_{\{t_2\le\phi_{i,m}(t_1)+0.5\}} - f_0$. Finally, define

$$\mathcal{F}_m \triangleq \{f_0 + \sum_{i=1}^{m} \zeta_i \psi_{i,m}(t_1,t_2), \quad \zeta_i \in \{0,1\}\}.$$

Since $\mathcal{F}_m \subset H^2(C)$, we have

$$\inf_{\hat{f}} \sup_{H^2(C)} \mathbb{E}\|f-\hat{f}\|_2^2 \ge \inf_{\hat{f}} \sup_{\mathcal{F}_m} \mathbb{E}\|f-\hat{f}\|_2^2. \tag{C.1}$$

The right hand side of (C.1) can be calculated more easily since we can restrict our attention to the estimators of the form $f_0 + \sum_{i=1}^{m} \hat{\zeta}_i \psi_{i,m}$. This is due to the fact that if $P_{\mathcal{F}_m}$ is the projection onto the above affine space, for every $f \in \mathcal{F}_m$ we have

$$\|P_{\mathcal{F}_m}(\hat{f}) - f\|_2^2 \le \|P_{\mathcal{F}_m}(\hat{f}) - P_{\mathcal{F}_m}(f)\|_2^2 \le \|\hat{f}-f\|_2^2.$$

Furthermore for any $f \in \mathcal{F}_m$ we have

$$\|f-\hat{f}\|_2 = \|\zeta - \hat{\zeta}\|_2,$$

and therefore the original problem reduces to one of estimating ζ. The final simplification is due to the fact that by projecting the observations onto the space of signals we can only improve the estimation. Therefore, we reduce the problem to the problem of estimating $\zeta_1,\ldots,\zeta_m$ from the observations $y_1^{\mathcal{F}},\ldots,y_m^{\mathcal{F}}$ given by

$$y_i^{\mathcal{F}} = \int \psi_{i,m}(t_1,t_2)dY(t_1,t_2) = \langle \psi_{i,m}, \psi_{i,m}\rangle \gamma_j + \int \psi_{j,m} dW(t_1,t_2).$$

Define $\langle \psi_{j,m}, \psi_{j,m}\rangle = \kappa_m$, we then have

$$y_i^{\mathcal{F}} = \kappa_m \zeta_i + w_i^{\mathcal{F}},$$

where $w_i^{\mathcal{F}} \overset{iid}{\sim} N(0,\sigma^2 k_m)$. Note that $\kappa_m = \Theta(m^{-3})$. Consider the problem of estimating γ_1 from the observation $\gamma_1 + N(0,\sigma^2/\kappa_m)$ and set m to the smallest integer for which $\sigma^2/\kappa_m \le 1$. If the risk of this estimator is lower bounded by B, the risk of the original estimator will be $Bm\kappa_m^2 = \Theta(\sigma^{4/3})$.

References

[1] L. I. Rudin, S. Osher, E. Fatemi, Physica D 60 (1992) 259–268.

[2] I. M. Johnstone, Gaussian processes: Sequence and multiresolution models, 2011. Unpublished.

[3] S. M. Smith, J. M. Brady, Int. J. of Computer Vision 23 (1997) 45–78.

[4] L. Yaroslavsky, Digital Image Processing – An Introduction, Springer Verlag, 1985.

[5] C. Tomasi, R. Manduchi, Proc. Int. Conf. Comp. Vision (1998) 839–846.

[6] M. Elad, IEEE Trans. Image Processing 11 (2002) 1141–1151.

[7] A. Buades, B. Coll, J. Morel, Proc. IEEE Int. Conf. Comp. Vision (2005).

[8] A. Maleki, M. Narayan, R. Baraniuk, Preprint (2011).

[9] D. L. Donoho, Ann. Stat. 27 (1999) 859 – 897.

[10] E. L. Lehman, G. Casella, Theory of Point Estimation, New York: Springer-Verlag, 1998.

[11] G. Peyrè, E. L. Pennec, C. Dossal, S. Mallat, in: Proc. SPIE, volume 6701.

[12] E. Candès, D. L. Donoho, Ann. Stat. 3 (2002) 784–842.

[13] A. Korostelev, A. Tsybakov, Minimax theory of Image Reconstruction, Lecture Notes in Statistics, Springer-Verlag, 1993.

[14] E. Arias-Castro, D. L. Donoho, Ann. Stat. 37 (2002) 1172–1206.

[15] E. J. Candès, D. L. Donoho, Philos. Trans. Roy. Soc. A 357 (1999) 2495–2509.

[16] E. Candès, D. L. Donoho, Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges, Technical Report, 1999.

[17] R. M. Willett, R. D. Nowak, IEEE Trans. Med. Imaging 22 (2003) 332 –350.

[18] G. Kutyniok, D. Labatte, J. Wavelet Theo. Appl. 1 (2007) 1–10.

[19] M. N. Do, M. Vetterli, IEEE Trans. Image Processing 14 (2005) 2091–2106.

[20] E. L. Pennec, S. Mallat, IEEE Trans. Image Processing 14 (2005) 423 –438.

[21] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, P. L. Dragotti, IEEE Trans. Image Processing (2006).

[22] A. Maleki, B. Rajaei, H. R. Pourreza, to appear in *IEEE Trans. Image Processing* (2011).

[23] X. Feng, P. Milanfar, in: Proc. Asilomar Conf. Signals, Systems, and Computers, volume 1, pp. 478–482.

[24] X. Li, M. Orchard, IEEE Trans. Image Processing 10 (2001) 813–817.

[25] J. Bigun, G. Granlund, J. Wiklund, IEEE Trans. Pattern Anal. Machine Intell. 13 (1991) 775–790.

[26] Barbara test image, `http://decsai.ugr.es/~javier/denoise/test_images/index.htm`, 2011.

[27] Boats test image, USC-SIPI Image Database, 2011.

[28] Wet-paint test image, `http://www.ece.rice.edu/~wakin/images/`, 2011.

[29] M. Nagao, T. Matsuyama, Computer Graphics and Image Processing 9 (1979) 394–407.

[30] H. Takeda, MSc thesis, University of California Santa Cruz (2006).

[31] R. Hummel, R. Moniot, IEEE Trans. Acoust., Speech, Signal Processing 37 (1989) 2111–2130.

[32] J. Koenderink, Biological Cybernetics 50 (1984).

[33] P. Perona, J. Malik, IEEE Trans. Pattern Anal. Machine Intell. 12 (1990) 629–639.

[34] J. W. Tukey, Exploratory Data Analysis, Addison Wesley, 1977.

[35] R. M. Willett, R. D. Nowak, IEEE Trans. Info. Theo. 53 (2007) 3171 –3187.

[36] M. Wakin, J. R. H. Choi, R. G. Baraniuk, IEEE Trans. on Image Processing 15 (2006) 1071–1087.

[37] A. Wong, P. Fieguth, D. Clausi, Proc. IEEE Int. Conf. Image Processing (ICIP) (2008) 537–540.

[38] C. Kervrann, J. Boulanger, IEEE Trans. Image Processing (2006) 2866–2878.

[39] M. Raphan, E. Simoncelli, An Empirical Bayesian Interpretation and Generalization of Nonlocal Means, Technical Report, Courant Institute of Mathematical Sciences, New York University, October 2010.

[40] A. Levin, B. Nadler, Computer Vision and Pattern Recognition (2011).

[41] P. Chatterjee, P. Milanfar, IEEE Trans. on Image Processing 19 (2010) 895–911.

[42] C. Stein, Approximate Computation of Expectation, Institute of Mathematical Statistics, 1986.