Suboptimality of Nonlocal Means on Images with Sharp Edges

Arian Maleki, Manjari Narayan, Richard Baraniuk

Department of Electrical and Computer Engineering
Rice University

Abstract—We conduct an asymptotic risk analysis of the nonlocal means image denoising algorithm for Horizon class images that are piecewise constant with a sharp edge discontinuity. We prove that the mean-square risk of nonlocal means is suboptimal and in fact is within a log factor of the mean square risk of wavelet thresholding.

I. INTRODUCTION

A. Denoising algorithms

The long history of image denoising (noise removal) is testimony to its central importance in image processing. A wide range of denoising schemes have been developed, ranging from simple convolutional smoothing and Wiener filtering to total variation methods [1] and sparsity-exploiting wavelet shrinkage [2]. One of the more successful denoising approaches proposed to date is the nonlocal means algorithm (NLM) [3]. NLM resembles convolutional smoothing in that it denoises a noisy image pixel value by replacing it with a weighted average of all of the pixel values in the image. However, in contrast to standard spatially invariant convolutional smoothing, which chooses the weights based on the spatial distances between pixels, NLM determines the weights according to the similarity of the neighborhoods surrounding the pixels.

The optimality of NLM is not completely understood.

NLM and related techniques have been stated to be optimal for denoising image textures [4], [5]. However, a theoretical understanding of its behavior on image edges is lacking. In this paper, we consider a prototypical model for images with edges, the so-called horizon model [6], [7]. Horizon class images consist of piecewise constant regions separated by a step edge discontinuity that follows a smooth contour in space. We calculate the asymptotic minimax risk of NLM estimator for horizon class images and show it to be suboptimal. Indeed, our results show that NLM has the same minimax risk as its competitor wavelet shrinkage. Our analysis has the side benefit of providing greater insight into the behavior of NLM.

B. Nonlocal means denoising

Consider a two-dimensional signal with "pixel values" \( x_{ij} \) for \( i, j \in \{1, 2, \ldots, n\} \) that is corrupted by iid Gaussian noise \( z_{ij} \sim N(0, \sigma^2) \). Denote the observations by

\[
y_{ij} = x_{ij} + z_{ij}, \quad \forall i, j \in \{1, 2, \ldots, n\}.
\]

Our goal is to estimate the noise-free values \( x_{ij} \) from the noisy observations \( y_{ij} \). Let \( \hat{f}_{ij} \) denote the estimate of \( x_{ij} \) for reasons that will be clear later. Following [3], we define the \( \delta_n \)-neighborhood distance \( d_{\delta_n}(y_{ij}, y_{m,\ell}) \) between two observations as

\[
d_{\delta_n}^2(y_{ij}, y_{m,\ell}) = \frac{1}{\rho_n^2} \sum_{m=-\delta_n}^{\delta_n} \sum_{\ell=-\delta_n}^{\delta_n} |y_{i+\ell,j+m} - y_{n+\ell,p+m}|^2 - |y_{ij} - y_{n,p}|^2,
\]

where \( \rho_n^2 = (2\delta_n + 1)^2 - 1 \). To simplify the calculations in Section III, we make one slight modification to the algorithm of [3] by removing the center element, \( |y_{ij} - y_{n,p}|^2 \), from the summation. Since \( \delta_n \to \infty \) as \( n \to \infty \), this removal will have a negligible effect on the asymptotic performance.

The nonlocal means algorithm uses the neighborhood distances to estimate

\[
\hat{f}_{ij}^N = \frac{\sum_{(m,\ell) \in S} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell) \in S} w_{m,\ell}},
\]

where \( S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \) and \( w_{m,\ell} \) is set according to the \( \delta_n \)-neighborhood distance between \( y_{ij} \) and \( y_{m,\ell} \). It is straightforward to verify that \( E(d_{\delta_n}^2(y_{ij}, y_{m,\ell})) = d_{\delta_n}^2(x_{ij}, x_{m,\ell}) + 2\sigma^2 \), which suggests the following strategy for setting the weights:

\[
w_{m,\ell} = \begin{cases} 
1 & \text{if } d_{\delta_n}^2(y_{ij}, y_{m,\ell}) \leq 2\sigma^2 + t_n, \\
0 & \text{otherwise},
\end{cases}
\]

where \( t_n \) is the threshold parameter. Soft/tapered versions of setting the weights have been explored and are often used [3], but the above untapered weights capture the main feature of the algorithm while simplifying our analysis. We introduce two oracle nonlocal means algorithms to facilitate our analysis. The full oracle NLM algorithm has access to \( E(d_{\delta_n}^2(y_{ij}, y_{m,\ell})) \) and thus sets the weights in (1) according to

\[
w_{m,\ell} = \begin{cases} 
1 & \text{if } d_{\delta_n}^2(x_{ij}, x_{m,\ell}) \leq t_n, \\
0 & \text{otherwise}.
\end{cases}
\]

Thesemi-oracle NLM algorithm (SNLM) only differs slightly from the standard NLM in that it uses the semi-oracle
neighborhood distance

\[ d_n^2(y_{i,j}, y_{n,p}) = \frac{1}{\rho_n^2} \sum_{m=-\delta_n}^{\delta_n} \sum_{\ell=-\delta_n}^{\delta_n} |x_{i+\ell,j+m} - y_{n+\ell,p+m}|^2 \]

\[ -\frac{1}{\rho_n^2} (x_{i,j} - y_{n,p})^2 \]

and then sets the weights in (1) according to

\[ w_{m,\ell} = \begin{cases} 1 & \text{if } d_n^2(y_{i,j}, y_{m,\ell}) \leq \sigma^2 + t_n, \\ 0 & \text{otherwise}. \end{cases} \]

Unlike the full-oracle, the semi-oracle assumes that just one-half of the noise is removed from the distance estimates. Therefore, the distances in SNLM are more accurate than the original nonlocal means but less precise than the full oracle. In the rest of the paper we will use \( f^N \), \( f^S \), and \( f^F \) to denote for the nonlocal means, semi-oracle nonlocal means, and full oracle nonlocal means algorithms, respectively.

C. Performance criterion

A particularly successful technique for analyzing the performance of a denoising algorithm is the non-parametric approach [6], [7]. Suppose we are interested in estimating a function \( f : [0, 1]^2 \to \mathbb{R} \) from noisy pixel-level observations. Define \( \text{Pixel}(i,j) = \left( \left[ \frac{i}{n}, \frac{i+1}{n} \right) \times \left[ \frac{j}{n}, \frac{j+1}{n} \right) \right) \) and let \( x_{i,j} = \text{Ave}(f | \text{Pixel}(i,j)) \) be the pixel-level averages of \( f \). We observe

\[ y_{i,j} = x_{i,j} + z_{i,j}, \]

where \( z_{i,j} \sim N(0, \sigma^2) \).

For a given function \( f \) and an estimator \( \hat{f} \) we define the risk function as

\[ R_n(f, \hat{f}) = \mathbb{E} \left( \frac{1}{n^2} \sum_i \sum_j (x_{i,j} - \hat{f}_{ij})^2 \right). \]

It can also be written as,

\[ R_n(f, \hat{f}) = \left( \frac{1}{n^2} \sum_i \sum_j (x_{i,j} - \mathbb{E}\hat{f}_{ij})^2 \right) \]

\[ + \mathbb{E} \left( \frac{1}{n^2} \sum_i \sum_j (\mathbb{E}\hat{f}_{ij} - \hat{f}_{ij})^2 \right), \]

where the first and second terms are called bias and variance of \( \hat{f} \), respectively. The risk defined in (6) depends on the specific choice of \( f \). When \( f \) is chosen from a class of functions \( \mathcal{F} \), we define the risk of the estimator on the class as

\[ R_n(\mathcal{F}, \hat{f}) = \sup_{f \in \mathcal{F}} R_n(f, \hat{f}). \]

In this paper we are interested in the asymptotic setting where \( n \) is very large. For the estimators we consider in this paper, \( R_n(\mathcal{F}, \hat{f}) \to 0 \) as \( n \to \infty \) at different rates. Therefore, the decay rate of the risk will be the performance measure in this paper. The minimax risk over functions in \( \mathcal{F} \) is then defined as the risk of the best estimator,

\[ R_n^*(\mathcal{F}) = \inf_{f \in \mathcal{F}} R_n(f, \hat{f}). \]

The minimax risk is a lower bound on the performance of all estimators for signals in \( \mathcal{F} \).

D. Horizon class edge model

Several different image edge models have been considered in the image processing and denoising literature. We use the Horizon class model that contains piecewise constant images containing sharp step edges [6], [7]. Let \( \text{H"older}^\alpha(C) \) be a class of H"older functions on \( \mathbb{R} \) defined as follows:

\[ h \in \text{H"older}^\alpha(C) \text{ if and only if } |h^{(k)}(x) - h^{(k)}(y)| \leq C|x-y|^{\alpha-k}, \]

where \( k = |\alpha| \). For a one-dimensional function \( h \), we define \( f_h : [0, 1]^2 \to \mathbb{R} \) as \( f_h(x, y) = 1_{h(x) < h(y)} \). Based on this mapping we define the Horizon class of signals as

\[ H^\alpha(C_1, C_\alpha) = \{ f_h(x, y) : h \in \text{H"older}^\alpha(C_\alpha) \cap \text{H"older}^1(C_1) \}. \]

In words, a horizon class image consists of two regions (one region taking the value 0 and the other taking the value 1) separated by a step discontinuity that runs along the curve \( h \). For the notational simplicity we assume that \( C_1 = 1 \). The following theorem specifies the minimax risk of the class of all measurable estimators on \( H^\alpha(C_1, C_\alpha) \).

\[ \text{Theorem 1.1: } [7] \text{ For } \alpha \geq 1, \text{ the minimax risk of the class } H^\alpha(C_1, C_\alpha) \text{ is } R_n^*(H^\alpha(C_1, C_\alpha)) = \Theta(n^{\frac{\alpha}{\alpha+2}}). \]

In [8], Donoho showed that isotropic wavelet thresholding estimator \( f^\text{wave} \) does not achieve the optimal minimax rate for \( \alpha > 1 \). In particular, he proved that \( \sup_{f \in H^\alpha(C_1, C_\alpha)} R_n(f, f^\text{wave}) = \Omega(\frac{1}{n}) \). This negative result spurred the development of more geometrical denoising techniques such as wedgelets [6] and curvelets [9]. However, while optimized for denoising image edges, these latter techniques performed poorly on image textures and other geometries.

II. MAIN RESULTS

Over the last decade, nonlocal means has emerged as a general-purpose denoising method of choice. A limited amount of performance analysis has been performed. Aujol et al. [10] have analyzed exemplar-based methods in general using geometric analysis on BV image models. Kerivann et al. and Raphan et al. [11], [12] have analyzed NLM from a Bayesian perspective. The goal of this paper is to perform a minimax risk analysis of the performance of NLM on Horizon class images with sharp edges. Our first theorem establishes an upper bound on the risk of NLM.
**Theorem 2.1:** Consider the NLM algorithm using \( \delta_n = \log n \) and \( t_n = \frac{2n^2}{\sqrt{\log n}} \). The risk of this algorithm over the class \( H^\alpha(C_1, C_\alpha) \) is

\[
\sup_{f \in H^\alpha(C_1, C_\alpha)} R_n(f; \hat{f}^N) = O \left( \frac{\log^3 n}{n} \right).
\]

This bound is within a \( \log^3(n) \) factor of the wavelet thresholding. However, this performance is suboptimal for \( \alpha > 1 \). In other words, NLM does not exploit the \( \alpha \)-smoothness of the edge contour.

The bound in Theorem 2.1 is for a specific choice of parameters, and hence it is natural to ask whether NLM can attain the optimal performance with some other choice of parameters. To answer this question, we consider SNLM, that outperforms standard NLM. We make the following mild assumptions:

- **NL1:** The window size \( \delta_n \to \infty \) as \( n \to \infty \). This assumption is critical to ensuring good performance of algorithm.
- **NL2:** The threshold is set to \( \sigma^2 + t_n \) as explained in (5), where \( t_n > 0 \). This assumption ensures that if the neighborhood of pixels around pixel \((m, \ell)\) is exactly the same as the neighborhood around the pixel \((i, j)\), then \( w_{m \ell} = 1 \) with high probability.
- **NL3:** The threshold \( t_n \) is set such that, if the noise-free neighborhoods are different in more than half of their pixels, i.e., if \( d^2(x_{i,j}, x_{m,\ell}) \geq \frac{1}{2} \), then

\[
P (w_{i,j}^2(m, \ell) = 1) = o \left( n^{-1} \right).
\]

**NL4:** \( \delta_n = O(n^\beta) \) for some \( \beta \leq 0.3 \).

Under these assumptions the following theorem provides a lower bound on the performance of NLM.

**Theorem 2.2:** Suppose that \( \delta_n \) and \( t_n \) satisfy NL1−NL4. The risk of the SNLM algorithm over the class \( H^\alpha(C_1, C_\alpha) \) satisfies

\[
\inf_{t_n, \delta_n} \sup_{f \in H^\alpha(C_1, C_\alpha)} R_n(f; \hat{f}^S) = \Omega \left( \frac{1}{n} \right).
\]

This upper bound is the same as that for wavelet thresholding, but still suboptimal compared to Theorem 1.1. To understand why NLM is suboptimal for Horizon class images, consider the particular Horizon class image displayed in Figure 1 and the estimation of an “edge” pixel. An edge pixel \((i, j)\) is a pixel that satisfies \( j = \lfloor nh(\frac{1}{2}) \rfloor \). Intuitively, we expect that a pixel below the edge has a non-zero weight with probability \( o(1) \). We prove that this probability is larger than \( p_0 \), where \( p_0 \) is independent of \( n \). Hence at most \( o(n) \) pixels below the edge can be included in the denoised estimate. Since \( x_{i,j} = 0 \), the bias will be larger than \( \frac{n p_0}{n + n p_0 + n p_0} \). Here \( n p_0 \) corresponds to the pixels below the edge that pass the threshold. Clearly the bias will be \( \Theta(1) \). This happens due to the low ‘signal to noise ratio’ in the patch distances near the edge. Combining this, with the fact that the distances are symmetric with respect to the edge, we prove that the risk on the edge elements is constant. The final step is to notice that there are \( O(n) \) edge pixels and therefore the risk over the whole image is \( \Omega \left( \frac{1}{n} \right) \).

III. THEORETICAL ANALYSIS

In this section we prove our main results. Section III-A includes the proof of Theorem 2.1 and section III-B summarizes the proof of Theorem 2.2.

A. Proof of Theorem 2.1

The proof has two main steps. (1) We show that the risk of the pixels far from the edge is \( O(\log^3 n/n) \). (2) We prove that the risk on the near edge pixels is constant. However, there are at most \( O(n \log n) \) of them.

We define the partitions as follows. Let \( S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \). For a given horizon function \( h(t_1, t_2) \), define

\[
S_1 = \{(i, j) \mid j > nh(\frac{1}{2}) + 2\delta_n\},
\]

\[
S_2 = \{(i, j) \mid nh(\frac{1}{2}) < j \leq h(\frac{1}{2}) + 2\delta_n\},
\]

\[
S_3 = \{(i, j) \mid nh(\frac{1}{2}) - 2\delta_n \leq j \leq nh(\frac{1}{2})\},
\]

\[
S_4 = \{(i, j) \mid j < nh(\frac{1}{2}) - 2\delta_n\}.
\]

These regions are displayed in Figure 2. We use the notation \( \sum_{(i,j) \in S_t} \) for a double summation over \( i, j \) where \( j \) satisfies the constraints specified for \( S_t \).

**Proof:** Consider a point \((i, j) \in S_1\). The risk of NLM at this pixel is

\[
\mathbb{E} \left( x_{i,j} - \frac{\sum w_{m \ell} y_{m \ell}}{\sum w_{m \ell}} \right)^2,
\]

where \( x_{i,j} = 0 \) since \((i, j) \in S_1\). Define the set of oracle weights

\[
w_{i,j}^* = \begin{cases} 1 & \text{if } \frac{j}{n} > h(\frac{1}{2}) \text{,} \\ 0 & \text{otherwise.} \end{cases}
\]

Call \( U = \left( \frac{\sum w_{m \ell} y_{m \ell}}{\sum w_{m \ell}} \right)^2 \), and let the event \( A = \{w_{m \ell} = \)}
The last inequality is the result of Cauchy-Schwartz. We bound the last three terms separately.

**Lemma 1:** Let \( w_{m,\ell} \) be the weights of NLM with \( \delta_n = \log n \) and \( t_n = \frac{1}{\sqrt{2\log n}} \) and let \( \omega^*_{m,\ell} \) be the oracle weights introduced in (9). Then

\[
E \left( \frac{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell}} \right)^2 
= O \left( \frac{\delta_n}{n} \right).
\]

**Proof:** We plug in the values of \( x_{i,j} \),

\[
E \left( \frac{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell}} \right)^2 
\leq E \left( \frac{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell}} \right)^2
= O \left( \frac{\delta_n}{n} \right),
\]

where the inequality (1) is due to the fact that it is an increasing function of \( \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell} \) and a decreasing function of \( \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell} \). Therefore, we set them accordingly.

**Lemma 2:** Let \( w_{m,\ell} \) be the weights of NLM with \( \delta_n = \log n \) and \( t_n = \frac{1}{\sqrt{2\log n}} \) and let \( w^*_{m,\ell} \) be the oracle weights introduced in (9). Then we have

\[
E \left( \frac{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell}} \right)^2
= O \left( \frac{\log^3 n}{n} \right).
\]

**Proof:** The proof of this lemma is long and therefore, we refer the interested reader to [13]. Here we just sketch the main steps. It is straightforward to see that we have to bound two terms, \( E(\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell})^2 \) and \( E(\sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell})^2 \). The first term is clearly bounded by \( O(1/n^2) \). We first write the second term as

\[
E \left( \frac{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} x_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w^*_{m,\ell} + \sum_{(m,\ell) \in S_{23}} w_{m,\ell}} \right)^2
\]

where \( C_{i,j}^n \) represents all the variables in the \( \delta_n \) neighborhood of \((i,j)\). We then use the conditional independencies and remove many terms from the summation to obtain the last equation. It is easy to confirm that the last summation is upper bounded by \( \log^3 n \).

Using Lemma 1 and Lemma 2 in (11) proves that

\[
E(U \mid A) P(A) = O \left( \frac{\delta^4}{n} \right).
\]

The final step of the proof for the case \((i, j) \in S_1\) is to prove an upper bound for \( P(A^c) \). The following lemma will be useful for this.

**Lemma 3:** Let \( Z \sim N(0, \sigma^2) \). For \( \lambda < \frac{1}{2\sigma^2} \), we have

\[
E(e^{\lambda Z^2}) = \frac{1}{\sqrt{1 - 2\lambda\sigma^2}}.
\]
Proof: The proof is a simple integral calculation:

\[
E(e^{\lambda Z^2}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(\lambda Z - \frac{Z^2}{2\sigma^2})} dZ = \frac{1}{\sigma\sqrt{1/2 - 2\lambda}}.
\]

Lemma 4: Let \( Z_1, Z_2, \ldots, Z_n \) be iid \( N(0, 1) \) random variables.

\[
P\left( \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - 1 > t \right) \leq e^{-\frac{t^2}{2}(1-\ln(1+t))}.
\]

Proof: Here we prove just the first claim; the proof of the second claim follows along very similar lines. From Markov’s Inequality, we have

\[
P\left( \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - 1 > t \right) \leq e^{-\lambda t - \lambda} E\left(e^{\lambda \frac{1}{n} \sum_{i=1}^{n} Z_i^2}\right) = e^{-\lambda t - \lambda} \left( e^{\lambda^2/n} \right)^n = e^{-\lambda t - \lambda} \left( 1 - \frac{2\lambda}{n} \right)^n.
\]

The last inequality follows from Lemma 3. The upper bound proved above holds for any \( \lambda < \frac{n}{2} \). To obtain the lowest upper bound we minimize \( e^{-\lambda t - \lambda} \left( 1 - \frac{2\lambda}{n} \right)^n \) over \( \lambda \). The optimal value of \( \lambda \) is \( \lambda^* = \frac{nt}{2(t+1)} \). Plugging \( \lambda^* \) into (13) proves the result.

Using this Lemma it is simple to prove that \( P(A^c) = O\left(\frac{1}{n^2}\right) \). See [13] for the details.

Combining the result of the above lemma with (12) we conclude that for \( (i,j) \in S_1 \) we have,

\[
E\left(x_{i,j} - \frac{\sum_{m \neq j} w_{m,j} y_{m,t}}{\sum_{m \neq j} w_{m,t}}\right)^2 = O\left(\frac{\log^3 n}{n}\right).
\]

Now consider \( (i,j) \in S_2 \cup S_3 \). In this region we just bound the error by 1. But, when we calculate the final risk we consider the number of these pixels to be on the order of \( O(n \log n) \). We combine the above results to obtain the desired upper bound.

Discussion of Theorem 2.1: From the proof provided above it is also clear that one can set the window size to \( \log^{(1+\varepsilon)/\varepsilon} n \) and the the threshold level to \( 2/\sqrt{\log^2 n} \) to obtain the upper bound \( O(\log^{(3+3\varepsilon)/2} N/n) \). Therefore, the performance of NLM may reach very close to \( O(\log^{1.5} n/n) \) in the asymptotic setting.

B. Proof of Theorem 2.2

In this section we focus on SNLM. We assume that the parameters satisfy assumptions NL1–NL4. Furthermore we consider the performance of the NLM on the image displayed in Figure 1. For the notational simplicity we assume that \( n \) is even. We first sketch the main steps of the proof. We just consider the edge pixels \( (j = \lfloor n/2 \rfloor) \) and prove that the risk at these pixels is a constant that does not depend on \( n \). The proof has the following steps.

- We prove that the probability a pixel just below the edge contributes to the ANLM estimate is larger than \( p_0 \), where \( p_0 \) is a non-zero probability independent of \( n \). See the formal statement in Theorem 1.
- We prove that the probability a pixel which is \( \ell < \delta_n/2 \) rows above the edge or below the edge contributes to the SNLM estimate is equal. This is formally stated in Lemma 6.
- We show that the risk will be minimal if all the edge pixels contribute to the estimate and the probability of the other pixels contributing is either zero of \( o(1/n) \). However, we have already proved that for \( \ell = 1 \) the probability is larger than \( p_0 \). Hence SNLM estimate is the average of the three rows and in Theorem 2.2 we show that the risk of this estimator is larger than a constant independent of \( n \).

Proposition 1: Let \( j^* = \lfloor \frac{n}{2} \rfloor \). For any pixel with coordinate of the form \((i^*,j^*)\) and for any value of \( \delta_n \) and \( t_n \), there exists a non-zero constant probability \( p_0 \) such that,

\[
P\left( \sum_{m} w_{m,j^*-1} - np_0 < -t \right) \leq 4\delta_ne^{-\frac{t^2}{2\pi}}.
\]

Proof: For the notational simplicity we use \( i = i^* \), \( j = j^* \), and \( f_{i,j} = f(\frac{i}{n}, \frac{j}{n}) \) in the proof.

\[
P(\overline{d}_{\rho}^2(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n) =
\]

\[
P\left( \frac{1}{\rho_n} \sum_{\ell} (s_{\ell,p}^2 - \sigma^2) - 2 \frac{1}{\rho_n} \sum_{\ell} s_{\ell,0} \leq -1 \frac{1}{\rho_n} + t_n \right) \geq
\]

\[
P\left( \frac{1}{\rho_n} \sum_{\ell} (s_{\ell,p}^2 - \sigma^2) - 2 \frac{1}{\rho_n} \sum_{\ell} s_{\ell,0} \leq -1 \frac{1}{\rho_n} \right),
\]

where \( s_{\ell,m} = z_{m+\ell,j-1}+p \). According to the Berry-Esseen Central Limit Theorem for independent non-identically distributed random variables, we know that,

\[
P\left( \frac{1}{\rho_n} \sum_{\ell} \sum_{p} (s_{\ell,p}^2 - \sigma^2) - 2 \frac{1}{\rho_n} \sum_{\ell} s_{\ell,0} \leq -1 \frac{1}{\rho_n} \right)
\]

\[
\geq P(G \leq -1) - \frac{C}{\rho_n},
\]

where \( G \) is a Gaussian random variable with mean zero and bounded standard deviation. In fact it is not difficult to confirm that,

\[
E(G^2) = 2\sigma^4 + \frac{8\sigma^2 \delta_n - 2\sigma^4}{(2\delta_n + 1)^2}.
\]

Since \( P(G \leq -1) \geq 2p_0 \) (2p_0 is \( P(G' \leq -1) \) where \( G' \sim N(0, 2\sigma^2) \)) is non-zero, for large values of \( n \) we can ensure that \( C/n < p_0 \) and therefore \( P(\overline{d}_{\rho}^2(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n) > p_0 \). We now prove that even though the weights are correlated, \( \Theta(n) \) of the weights will be equal to 1 with very high probability. Call \( u_i = w_{i,j-1} \) and define the
process \( U = (u_1, \ldots, u_n) \). We break this sequence into \( 2\delta_n \) subsequences \( U_i = (u_{i}, u_{i}+\delta_n, \ldots, u_{i}+2\delta_n) \). Each \( U_i \) has independent and identically distributed elements. Therefore according to the Hoeffding inequality we have,
\[
P(\sum_{u_j \in U_i} u_j - \frac{n}{2\delta_n} p_0 > t) \leq 2e^{-\frac{2t^2}{n}}.
\]

Based on this, we have
\[
P\left( \sum_{i \in U_i} u_j - \frac{n}{2\delta_n} p_0 - t \right) \leq 4\delta_n e^{-\frac{t^2}{n\delta_n}}.
\]

and the result of Theorem 1 follows.

We can summarize the result of the above theorem in the following Corollary.

**Corollary 1:** Let \( \delta_n = O(n^\alpha) \) for \( \alpha < 1 \). For any \( \delta_n \) and \( t_n \), with very high probability \( \Theta(n) \) of the pixels just below the edge will pass the threshold.

**Proof:** Set \( t = n^{\frac{2\alpha}{1+\alpha}} \) in the previous theorem. 

Remarkably the above corollary holds in a very general setting even if the assumptions \( NL_1 \sim NL_4 \) do not hold. In other words NLM in its most general form is not able to discriminate among the pixels that are very close to the edge. This is due to the fact that the signal to noise ratio in the similarity estimates is pretty low at the edge pixels.

**Lemma 5:** If \( |m - i^*| > \delta_n/2 \) and \( |m' - i^*| > \delta_n/2 \),
\[
P(\delta_{\alpha_n}^2 (y_{i^*, j^*}, y_{m, j^*} - \ell) \leq \sigma^2 + t_n)
= P(\delta_{\alpha_n}^2 (y_{i^*, j^*}, y_{m', j^*} - \ell) \leq \sigma^2 + t_n).
\]

for any \( \ell, m, m' \).

**Lemma 6:** For \( \ell < \delta_n/2 \),
\[
P(\delta_{\alpha_n}^2 (y_{i^*, j^*}, y_{m, j^*} - \ell) \leq \sigma^2 + t_n)
= P(\delta_{\alpha_n}^2 (y_{i^*, j^*}, y_{m, j^*} + \ell) \leq \sigma^2 + t_n).
\]

The proof of this lemma is obvious and is skipped here.

**Theorem 2.2** Let the SNLM satisfy \( NL_1 \sim NL_4 \). Then
\[
\inf_{t_n, \delta_n} \sup_{f \in H^\infty(C)} R_n(f, f^S) = \Omega\left(\frac{1}{n}\right).
\]

**Proof:** We derive a lower bound for the risk of SNLM on the image displayed in Figure 1. To do so, we consider the pixels just above the edge and prove that the SNLM algorithm has risk \( \Theta(1) \) at these pixels. Since there are \( \Theta(n) \) of these pixels, the risk over the entire image is larger than \( \Theta(n^{-1}) \).

Consider a pixel \((i^*, j^*) \) with \( j^* = \left[ \frac{n}{2} \right] \). The risk of the SNLM is
\[
E \left( \sum_{m, \ell} \frac{w_{m, \ell} y_{m, \ell}}{\sum w_{m, \ell}} \right)^2 \geq \left( \sum_{m, \ell} \frac{w_{m, \ell} y_{m, \ell}}{\sum w_{m, \ell}} \right)^2.
\]

Note that \( w_{m, \ell} \) is independent of the \( y_{m, \ell} \) according to the construction of the NL weights in (4). Let \( p_{n, \ell} \) be the probability \( P(w_{m, \ell} = 1) \) for \( \ell \in \{ j^* - \delta_n, j^* - \delta_n + 1, \ldots, j^* + \delta_n \} \). We can partition the row \( \{ (i, \ell) \mid 1 \leq i \leq n \} \) into \( 2\delta_n + 1 \) subsequences and apply Hoeffding inequality on each subsequence. We combine the results of different subsequences with the union bound to prove that
\[
P\left( \sum_{m} w_{m, \ell} - n p_{n, \ell} > t \right) \leq 4\delta_n e^{-\frac{t^2}{n\delta_n}}.
\]

Define the event \( A \) as
\[
A = \left\{ \sum_{m} w_{m, \ell} - n p_{n, \ell} < n^{0.66} \forall \ell, |\ell - j^*| \leq \delta_n \right\}.
\]

Using the union bound and (15) we have
\[
P(A^c) \leq 8\delta_n e^{-\frac{t^2}{n\delta_n}}.
\]

Any lower bound on the bias of the estimator leads to a lower bound on its risk. Therefore, we find a lower bound for the bias as follows:
\[
E \left( \sum_{m, \ell} \frac{w_{m, \ell} y_{m, \ell}}{\sum w_{m, \ell}} \right) \geq \left( \sum_{m, \ell} \frac{w_{m, \ell} y_{m, \ell}}{\sum w_{m, \ell}} \right) - P(A^c),
\]

where for the last inequality we have used the fact that the risk of SNLM is bounded by 1. Since from the construction of SNLM in (4), \( w_{m, \ell} \) is independent of \( z_{m, \ell} \), we have
\[
E \left( \sum_{m, \ell} \frac{w_{m, \ell} y_{m, \ell}}{\sum n p_{n, \ell} + n^{0.66} \delta_n} \right) - P(A^c)
= \sum_{m, \ell} \left( \frac{w_{m, \ell} y_{m, \ell}}{\sum n p_{n, \ell} + n^{0.66} \delta_n} \right) - P(A^c)
\geq \sum_{m, \ell} \left( \frac{n p_{n, \ell} + n^{0.66} \delta_n}{\sum n p_{n, \ell} + n^{0.66} \delta_n} \right) - P(A^c).
\]
fact with Lemma 6, we obtain
\[
\sum_{\ell<j^*} n p_{n,\ell} \geq \frac{n p_{0} + n + n^6 \delta_n}{p_{0} + 1 (1 + o(1))}.
\]

This completes the proof.

**Discussion of Theorem 2.2:** From Theorem 2.2 it is clear that the NLM is not able to exploit the smoothness of the edges to improve the performance of the denoiser and therefore except for \( \alpha = 1 \) the decay it provides for the horizon class is suboptimal.

**IV. CONCLUSION**

In this paper we have provided what is in the best of our knowledge the first asymptotic risk analysis of the NLM algorithm for images with sharp edges. We proved that NLM is blind to the smoothness of edges along the edge contour, and hence it does not offer optimal performance. In fact, the performance of NLM is within a log factor of that of wavelet thresholding.

**REFERENCES**

[1] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D.*, 60(1-4):259–268, 1992.

[2] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Stat.*, 26(3):879–921, Jan 1998.

[3] A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *SIAM J. Multiscale Model. and Sim.*, 4(2):490–530, Jan 2005.

[4] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. pages 1033–1038, 1999.

[5] E. Levina and P. J. Bickel. Texture synthesis and nonparametric resampling of random fields. *Ann. Stat.*, 34(4):1751–1773, 2006.

[6] D. L. Donoho. Wedgelets: Nearly minimax estimation of edges. *Ann. Stat.*, 27(3):859 – 897, Jan 1999.

[7] A.P. Korostelev and A.B. Tsybakov. *Minimax theory of Image Reconstruction*. Lecture Notes in Statistics. Springer-Verlag, 1993.

[8] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Info. Theo.*, 9(9):1532–1546, Jan 1995.

[9] Emmanuel Candes and David L. Donoho. Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges. Technical report, 1999.

[10] J. Aujol and S. Ladjal. Exemplar-based inpainting from a variational point of view. Preprint, Jan 2009.

[11] C. Kervrann, J. Boulanger, and P. Coupe. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. *Lect. Notes Comp. Sci.*, Jan 2007.

[12] M. Raphan and E. Simoncelli. An empirical bayesian interpretation and generalization of nonlocal means. Technical report, Courant Institute of Mathematical Sciences, New York University, October 2010.

[13] A. Maleki, M. Narayan, and R. Baraniuk. Minimax analysis of the non-local means algorithm. Preprint, 2011.