A Context Quantization Approach to Universal Denoising

Kamakshi Sivaramakrishnan
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
ksivaram@stanford.edu

Tsachy Weissman
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
tsachy@stanford.edu

***Abstract*— We revisit the problem of denoising a discrete-time continuous-amplitude signal corrupted by a known memoryless channel. By modifying our earlier approach to the problem, we obtain schemes that are much more tractable than the original ones, while retaining their universal optimality properties. The schemes involve a simple preprocessing step of quantizing the noisy symbols to generate quantized *contexts* which (according to the quantized context value of each symbol) are then used to partition the unquantized symbols to subsequences. A universal context-free denoiser (of zero context length for the unquantized sequences) is then separately employed on each of the subsequences. We identify a rate in which the context length and quantization resolution should be increased so that the resulting scheme is universal in both the semi-stochastic and fully stochastic settings. The proposed family of schemes is computationally attractive, having linear complexity with a proportionality constant that is independent of the context length and the quantization resolution. Experimental results show that these schemes are not only superior from a computational viewpoint, but also achieve better denoising in practice.**

I. Introduction

Consider the problem of estimating the clean signal $\{X_t\}_{t\in\mathbb{T}}$, $X_t \in [a,b] \subset \mathbb{R}$, from its noisy observations $\{Y_t\}_{t\in\mathbb{T}}$, $Y_t \in \mathbb{R}$, where $\{Y_t\}$ is the output of a known, memoryless channel whose input is $\{X_t\}$. Recently, universal denoising for continuous-valued signals and channels was considered in [1], [2], [3], [4]. This approach was motivated by the DUDE framework in [5], [6] and nonparametric techniques in density estimation [7]. This framework is a "two-pass" approach in which the first pass involves accruing the statistics of the noisy sequence and using knowledge of the channel to accrue statistics of the underlying clean signal itself. The second pass is one where, having learned the statistics, denoising is carried out to minimize a user-specified loss function, which penalizes incorrect guesses of the underlying clean signal. The denoising approach in [1] addresses the problem of sparse statistics that affects its discrete counterpart in [5] using natural context aggregation by resorting to nonparametric density estimation techniques. Universal optimality of the proposed denoisers in [1], [2] was established in a generality that applies to arbitrarily distributed clean signals, arbitrary memoryless channels and loss functions (with some benign regularity conditions). The theoretical optimality results were also translated to some encouraging practically implementable schemes discussed in [2], [3]. However, this family of denoisers suffers from computational issues that could render the schemes practically unattractive.

The context-free denoiser discussed in [2], where it is called the symbol-by-symbol denoiser, is computationally very attractive. However, greater advantage of this denoiser is gained only by considering higher context lengths of the sliding window denoiser which comes with increased computational burden. One possible approach to reduce the computational burden could be to quantize the continuous-valued noisy symbols (as proposed in [6] for the discrete-input general output setting) and then apply higher order sliding window denoisers. This is going to necessarily do worse than learning the statistics of the noisy sequence from the *true* unquantized values. In this paper we propose a middle ground solution, where for any order of the denoiser, we quantize the *contexts* of every noisy symbol and apply the context-free denoiser of [2] to each class of quantized contexts. This approach emulates the higher order functionality of the sliding window denoisers through the quantized contexts while still maintaining the low complexity of the context-free denoiser of [2] as it is applied within each context subsequence. The complexity of the denoisers obtained in this way is not only linear in the data size n for a fixed context length k and quantization resolution M, but in fact is bounded with a proportionality constant that does not depend on k and M. This is in stark contrast to the complexity of the scheme in [2] which, although linear in n, is exponential in k and is, consequently, impractical for even moderate values of k. The question is whether the new denoiser we propose, beyond its superiority over that of [2] from a computational standpoint, preserves the asymptotic universal optimality properties of the denoiser in [2]. In this paper, we answer this question in the affirmative.

The remainder of the paper is organized as follows. In section II, we discuss the problem setup and notations. This is followed by a brief discussion of the denoiser in [2] and some key technical results therein. Section IV details the construction of the proposed denoiser and performance guarantees for our suggested universal denoiser in the semi-stochastic setting. The performance guarantee is proved by comparing the proposed denoiser to the minimum possible "symbol-by-symbol" loss incurred by a denoiser that chooses from a set of experts that also make decisions based on quantized contexts. We will be extending these results, with diminishing quantization step-sizes, to eventually compare our performance to that of the denoiser in [2] which would establish the asymptotic optimality. Section V briefly mentions some promising preliminary experimental results.

II. Problem Setting and Notations

Similar to the problem setting in [1], let $\mathbf{x} = (x_1, x_2, \cdots)$ be the individual noise-free source signal with components taking values in $[a, b] \subset \mathbb{R}$ and $\mathbf{Y} = (Y_1, Y_2, \cdots)$, $Y_i \in \mathbb{R}$ be the corresponding noisy observations, also referred to as the output of the channel (corruption source). The channel considered here is memoryless, specified by a family of distribution functions $\mathcal{C} = \{F_{Y|x}\}_{x \in [a,b]}$, where $F_{Y|x}$ denotes the distribution of the channel output symbol when the input symbol is x. We assume the associated family of measures $\Upsilon = \{\mu_x\}_{x \in [a,b]}$ to be tight in the sense that $\sup_{x \in [a,b]} \mu_x([-T, T]^c) \to 0$ as $T \to \infty$. We make some additional assumptions on the nature of the channel which are elaborated in detail in [2]

An *n-block denoiser* is a measurable mapping taking $\mathbb{R}^n$ into $[a, b]^n$. We assume a loss function $\Lambda : [a, b]^2 \to [0, \infty)$ and denote the normalized cumulative loss of an n-block denoiser $\hat{X}^n$ by

$$L_{\hat{X}^n}(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} \Lambda(x_i, \hat{X}^n(y^n)[i]) \tag{1}$$

where $\hat{X}^n(y^n)[i]$ denotes the i-th component of $\hat{X}^n(y^n)$. We denote $\Lambda_{\max} = \sup_{x,y \in [a,b]} \Lambda(x, y)$, and assume $\Lambda_{\max} < \infty$. We also impose mild regularity conditions on the permissible loss functions, Λ, which are detailed, again, in [2]. Define the symbol-by-symbol minimum loss of x^n by

$$D_0(x^n) = \min_g E\left[\frac{1}{n} \sum_{i=1}^{n} \Lambda(x_i, g(Y_i))\right] \tag{2}$$

where the minimum is over all measurable maps $g : \mathbb{R} \to [a, b]$. For $x^n \in [a, b]^n$ define

$$F_{x^n}(x) = \frac{|\{1 \le i \le n : x_i \le x\}|}{n}, \tag{3}$$

i.e., the CDF associated with the empirical distribution of x^n. It is shown in [1], the minimizer of (2) is achieved by $g_{\text{opt}}[F_{x^n}]$ where, g_{opt} is given by

$$g_{opt}[F](y) = \arg\min_{\hat{x} \in [a,b]} \int_{[a,b]} \Lambda(x, \hat{x}) f_{Y|x}(y) dF(x) \tag{4}$$

III. Universal Denoising of Continuous-Amplitude data

In this section, we briefly recap the construction of the denoiser discussed in detail in [2],[1]. We also state the important result bounding the deviation of the loss incurred by our proposed denoiser from minimum possible symbol-by-symbol loss, $D(x^n)$. We refer the reader to [2] for further details and proofs of the result.

A. Construction of the Denoiser

F_{x^n} and, hence, $g_{opt}[F_{x^n}]$ are not known to an observer of the noisy sequence, Y^n. For an input sequence x^n, given the memoryless nature of the channel, the output symbols will be distributed as $\{F_{Y|x_1}, \cdots, F_{Y|x_n}\}$ and have the corresponding density functions, $\{f_{Y|x_1}, \cdots, f_{Y|x_n}\}$. Given the memoryless nature of the channel, the sequence of output symbols, $Y_1, Y_2, \cdots, Y_n$ are independent random variables taking values in $\mathbb{R}$ and have conditional densities, $f_{Y|x_1}, f_{Y|x_2}, \cdots, f_{Y|x_n}$ respectively. A density estimate is a sequence $f^1, f^2, \cdots, f^n$, where for each n, $f_Y^n(y) = f^n(y; Y_1, \cdots, Y_n)$ is a real-valued Borel measurable function of its arguments, and for fixed n, f_Y^n is a density estimate on $\mathbb{R}$. The *kernel estimate* is given by

$$f_Y^n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right) \tag{5}$$

where $h = h_n$ is a sequence of positive numbers and K is a Borel measurable function satisfying $K \ge 0$, $\int K = 1$.

Once we have an estimate $f_Y^n = f_Y^n[Y^n]$ for this function, we use it to estimate the input empirical distribution by

$$\hat{F}_{x^n}[Y^n] = \arg\min_{F \in \mathcal{F}_n^{[a,b]}} d\left(f_Y^n, \underbrace{\int f_{Y|x} dF(x)}_{[F \otimes \mathcal{C}]_Y}\right) \tag{6}$$

where $\mathcal{F}_n^{[a,b]} \subseteq \mathcal{F}^{[a,b]}$ denotes the set of empirical distributions induced by n-tuples with $[a, b]$-valued components and $[F \otimes \mathcal{C}]_Y$ denotes the marginal density induced at the output of the channel by an input distribution F. The definition for the norm, d, is

$$d(f, g) = \int |f(y) - g(y)| \, dy \tag{7}$$

A two-stage quantization of both, the support of the underlying clean symbol, $[a, b]$, and the levels of the estimate of its empirical distribution function, $\hat{F}_{x^n}$, itself is carried out to give the corresponding quantized probability mass function that has mass points only at the quantized symbols. The quantization of the interval $[a, b]$ is depicted in Fig. 1 below.

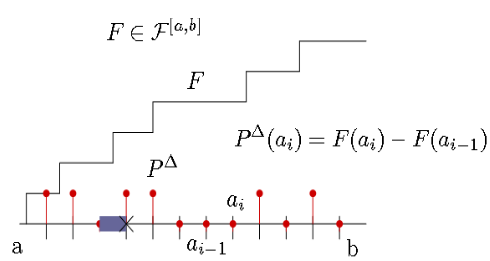

Fig. 1. Quantization of the support of a distribution function, $F \in \mathcal{F}^{[a,b]}$, $\mathcal{A}^{\Delta} = \{a_i = a + i\Delta, i = 0, \cdots, N(\Delta) - 1\}$, $N(\Delta)$ is the number of points in the interval $[a, b]$ corresponding to a quantization step-size of Δ

Applying this quantization of the support of the underlying clean symbol to the estimate, $\hat{F}_{x^n}$, we construct now, the corresponding probability mass function, $\hat{P}^{\Delta}_{x^n}$

$$\hat{P}^{\Delta}_{x^n}(a_i) = \hat{F}_{x^n}(a_i) - \hat{F}_{x^n}(a_{i-1}) \tag{8}$$

where, $a_i \in \mathcal{A}^{\Delta}$. The quantization of the values $\hat{P}_{x^n}$ is carried out using a uniform quantizer, Q_δ

$$\tilde{P}^{\delta,\Delta}_{x^n} = Q_\delta(\hat{P}^{\Delta}_{x^n}) \tag{9}$$

The minimizer of the Bayes envelope in (4) is then constructed from the quantized probability mass function, $P^{\delta,\Delta}_{x^n}$,as $g_{\text{opt}}\left[P^{\delta,\Delta}_{x^n}\right]$, where g_{opt} for the quantized clean symbol is,

$$g_{\text{opt}}[P](y) = \arg\min_{\hat{x}\in\mathcal{A}^{\Delta}} \sum_{a\in\mathcal{A}^{\Delta}} \Lambda\left(a,\hat{x}\right)\cdot f_{Y|x=a}(y)\cdot P\left(X=a\right) \tag{10}$$

Equipped with $P^{\delta,\Delta}_{x^n}$, the n-block context-free denoiser is now given by,

$$\tilde{X}^{n,\delta,\Delta}[y^n](i) = g_{\text{opt}}[\tilde{P}^{\delta,\Delta}_{x^n}[y^n]](y_i), \qquad 1\le i\le n \tag{11}$$

where, g_{opt} is given in (10).

The extension of the symbol-by-symbol scheme in (11) to the k^{th}-order sliding window case is carried out by subsequencing [1],[2] and is given by,

$$\tilde{X}^{n,\delta,\Delta,k} = \{\tilde{X}^{n_i,\delta,\Delta,k}\}_{1\le i\le 2k+1} \tag{12}$$

where, the denoiser for each of the subsequences, i, is

$$\tilde{X}^{n_i,\delta,\Delta,k}[y^n](j) = g_{\text{opt}}\left[\tilde{P}^{\delta,\Delta,k}_{x^{n_i}}[y^{n_i}]\right]\left(y^{j+k}_{j-k}\right), \\ j\in\{k+i, 3k+1+i,\cdots\lceil\frac{n-2k-i-1}{2k+1}\rceil\} \tag{13}$$

with $\tilde{P}^{\delta,\Delta,k}_{x^n}$ being the k^{th}-order equivalent of $P^{\delta,\Delta}_{x^n}$ and

$$g_{\text{opt}}[P]\left(y^k_{-k}\right) = \arg\min_{\hat{x}\in\mathcal{A}} \Lambda(\cdot,\hat{x})^T[P\otimes\mathcal{C}]_{X|y^k_{-k}} \\ = \arg\min_{\hat{x}\in\mathcal{A}} \sum_{\hat{a}\in\mathcal{A}} \Lambda\left(a,\hat{x}\right)\cdot\left\{\sum_{x^k_{-k}\in\mathcal{A}^{2k+1}:x_0=a}\left[\prod_{i=-k}^{k} f_{Y|x=x_i}(y_i)\right.\right. \\ \left.\left. P\left(X^k_{-k}=x^k_{-k}\right)\right]\right\} \tag{14}$$

B. Analysis

We state the important result derived in detail in [2] that, for a given sequence x^n, bounds the deviation of the cumulative incurred by the denoiser in (11) from the minimum possible symbol-by-symbol loss, $D_0(x^n)$. An important consequence of this result is Lemma 1 in section IV-B which gives a similar result for the proposed Modified Denoiser.

Theorem 1: For all $\epsilon > 0$, $\delta > 0$, $\Delta > 0$ there exist $\zeta = \zeta\left(\mathcal{C},\lambda,\Delta\right)$, $\psi = \psi\left(\mathcal{C},\epsilon,\delta,\Delta,\Lambda,\Delta\right)$, $A = A\left(\Lambda,\epsilon,\delta,\Delta\right)$ $n_0\left(\mathcal{C},K,\{h\},\epsilon,\Delta,\Lambda\right)$ s.t.

$$P\left(|L_{\tilde{X}^{n,\delta,\Delta}}(x^n,Y^n) - D_0(x^n)| > \epsilon+\delta\Lambda_{\max}+\zeta\right) \\ \le Ae^{-n\psi}, \qquad \forall n > n_0\left(\mathcal{C},K,\{h\},\epsilon,\Delta,\Lambda\right) \text{ and } x^n\in[a,b]^n$$

where, and $\lim_{\Delta\to 0}\zeta\left(\mathcal{C},\Lambda,\Delta\right) = 0$. The precise functional forms of ζ, ψ and χ can be inferred from Theorem 4 in [2]. Analogous result is derived for the k^{th}-order sliding window denoiser in [2].

IV. The Modified Universal Denoiser

The k^{th}-order sliding window denoiser discussed in section III is in principle an elegant approach to the problem of universal denoising of continuous-amplitude data. This denoiser systematically approaches the problem using nonparametric techniques to learn a quantized version of the aposteriori distribution of the underlying clean symbol given its observed noisy context, $[\tilde{P}^{\delta,\Delta}_{x^n}\otimes\mathcal{C}]_{X|y^k_{-k}}$ induced by the channel $\mathcal{C}$. As discussed in the introduction this scheme is, however, computationally intensive with increasing context lengths, k. Motivated, primarily, by the need to reduce computational burden of the scheme in section III, we propose and study the modification shown in Fig. 2. In words, we are proposing the following,

Fix, a window size k and number of levels of quantization, M

- Quantize the $2k$-length contexts,$\left(y^{-1}_{-k}, y^k_1\right)$, using an M-level vector quantizer to give M possible $2k$-length tuples, $\left(\hat{y}^{-1}_{-k},\hat{y}^k_1\right) \in \mathbf{Y}^M$. For a given M, $\mathbf{Y}^M = \{\mathbf{Y}_1,\cdots,\mathbf{Y}_m,\cdots,\mathbf{Y}_M\}$ denotes the sequence of M, $2k$-tuples in which the quantized contexts, $\left(\hat{y}^{-1}_{-k},\hat{y}^k_1\right)$, can take values. Let n_m be the number of $2k$-tuples at quantization level, m.
- For each, m, collect all the *unquantized* middle symbols that have quantized contexts, $\left(\hat{y}^{-1}_{-k},\hat{y}^k_1\right) = \mathbf{Y}_m$ to form $y^{n,m}_0 = \left\{y_j : \left(\hat{y}^{j-1}_{j-k},\hat{y}^{j+k}_{j+1}\right) = \mathbf{Y}_m\right\}$. Apply the symbol-by-symbol scheme discussed in section III to this collection of middle symbols. This corresponds to a bank of M context-free denoisers as shown in Fig. 2.
- Collect the denoised estimates $\hat{x}^{n,m}_0$, (*in appropriate order*) from all the quantization levels, $m\in\{1,\cdots,M\}$ to give the denoised sequence $\hat{x}^n$.

The specifics of the construction of this Modified Denoiser is discussed in the following subsection.

A. Construction of the Denoiser

The application of the density estimator in (5) at each of the M branches of the *Modified Denoiser* gives an estimate of the quantity, $f^n_{Y_0|\mathbf{Y}_m}(y)$. Thus, for a quantization level, m, (5) becomes

$$f^n_{Y_0|\mathbf{Y}_m}(y) = \frac{1}{n_m h}\sum_{j:\left(\hat{Y}^{-1}_{j-k},\hat{Y}^{j+k}_1\right)=\mathbf{Y}_m} K\left(\frac{y-Y_j}{h}\right), \\ m\in 1,\cdots,M \tag{15}$$

where, Y_j are such that $\left(\hat{Y}^{j-1}_{j-k},\hat{Y}^{j+k}_{j+1}\right) = \mathbf{Y}_m$, the quantized $2k$-tuple at level m. Additionally, application of (6) to $f^n_{Y_0|\mathbf{Y}_m}(y)$ now gives

$$\hat{F}^m_{x_0|\mathbf{Y}_m}[Y^{n,m}_0] = \arg\min_{F\in\mathcal{F}^{[a,b]}_{n_m}} d\left(f^n_{Y_0|\mathbf{Y}_m}, \underbrace{\int f_{Y|x}dF(x)}_{[F\otimes\mathcal{C}]_Y}\right) \tag{16}$$

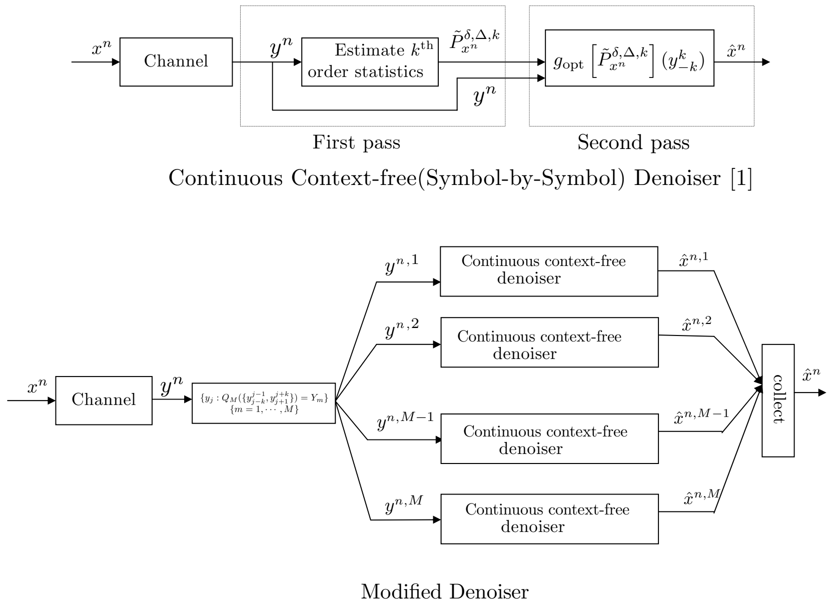

Fig. 2. Proposed modification to the Denoiser of [1]

which is an estimate of $F^m_{x_0|\mathbf{Y}_m}$, the *true* empirical distribution of the underlying clean symbol given the observed noisy context quantized to level m. Finally, the two-stage quantization for step-sizes, δ and Δ, corresponding to the support of the underlying clean symbol, $[a,b]$ and distribution function levels respectively, gives $\tilde{P}^{m,\delta,\Delta}_{x_0|\mathbf{Y}_m}$. For a given M, k, δ, Δ the denoiser candidate is now given by the sequence,

$$\tilde{X}^{n,M,\delta,\Delta,k} = \{\tilde{X}^{n_m,\delta,\Delta,k}\}_{1\leq m\leq M} \quad (17)$$

where

$$\tilde{X}^{n_m,\delta,\Delta,k}[y^{n,m}](j) = g_{\text{opt}}[\tilde{P}^{m,\delta,\Delta}_{x_0|\mathbf{Y}_m}[y^{n,m}]](y^{n,m}_j)$$

$y^{n,m} = \{y^{n,m}_j\}, j = 1,\cdots,n_m$ and g_{opt} is given by (10). The cumulative loss of the proposed denoiser is then given by,

$$L_{\tilde{X}^{n,M,\delta,\Delta,k}}(x^n, Y^n) = \sum_{m=1}^{M} L_{\tilde{X}^{n_m,\delta,\Delta,k}}(x^{n,m}, Y^{n,m}) \quad (18)$$

where, x^{n_m} is the underlying clean sequence corresponding to y^{n_m}.

B. Analysis

Let, $D^M_k(x^n)$ be the minimum possible k^{th}-order sliding window loss corresponding to an underlying clean sequence, x^n and noisy sequence Y^n with M-level quantization of noisy contexts. Specifically,

$$D^M_k(x^n) = \min_g E\left[\frac{1}{n-2k}\sum_{i=k+1}^{n-k}\Lambda(x_i, g(\hat{Y}^{i-1}_{i-k}, \hat{Y}^{i+k}_{i+1}, Y_i))\right] \quad (19)$$

where, $g : \mathbf{Y}^M \times \mathbb{R} \to \mathbb{R}$,

$$\left(\hat{Y}^{-1}_{-k}, \hat{Y}^k_1\right) = Q_M\left(\left(Y^{-1}_{-k}, Y^k_1\right)\right) \quad (20)$$

with, $Q_M : \mathbb{R}^{2k} \to \mathbf{Y}^M$. Simultaneously, we also define D^m_k which is the same as (19) except the minimum is over all $g : \mathbf{Y}_m \times \mathbb{R} \to \mathbb{R}$, $m \in 1,\cdots,M$. In other words, D^m_k is the minimum possible k^{th}-order sliding window loss corresponding to an underlying clean sequence, $x^{n,m}$ and (the corresponding) noisy sequence $Y^{n,m}$. From the construction of the denoiser, note that

$$D^m_k(x^n) = D_0(x^{n,m})$$

Using the result in Theorem 1 we can state the following Lemma which, for a sequence x^n, bounds the deviation of the cumulative loss incurred by the proposed denoiser at quantization level m from $D^m_k(x^{n,m})$,

Lemma 1: For all $\epsilon > 0$, $\delta > 0$, $\Delta > 0$, $1 \leq k \leq \lfloor \frac{n}{2} \rfloor$, at each quantization level, $m \in \{1,\cdots,M\}$, $\exists\ \zeta^k$, ψ^k, A^k, n^m_0 s.t.

$$P\left(|L_{\tilde{X}^{n_m,\delta,\Delta,k}}(x^{n,m}, Y^{n,m}) - D_0(x^{n,m})| > \epsilon + \delta\Lambda_{\max} + \zeta^k\right) \leq A^k e^{-n_m\psi^k},\ \forall n_m > n^m_0 \text{ and } x^{n,m} \in [a,b]^{n_m}$$

where, ζ^k, ψ^k, A^k are k^{th}-order equivalents of ζ, ψ, A (also functions of $\mathcal{C}, \delta, \Delta$, Λ and ϵ) in Theorem 1 whose precise forms can again be deduced from Theorem 12 in [2]. n^m_0 has the same functional form as the n_0 in Theorem 1 but, also depends on the quantization level m through the induced distribution of the quantized noisy symbols and k through the conditional distribution, $F_{x_0|\mathbf{Y}_m}$, induced by the k^{th}-order contexts.

The following Lemma formalizes the fact that, for any sequence, by performing *optimally* within every quantized context, wherein we allow the denoiser candidate to be different for every quantization level, we will be doing at least as well as the scheme which fixes one denoiser for all the quantization levels.

Lemma 2: For any sequence, x^n and $M > 0$

$$D_k^M(x^n) \geq \frac{1}{M}\sum_{m=1}^{M} D_k^m(x^n) \tag{21}$$

Using this fact, we state the following theorem which, for a given number of quantization levels, M of the noisy contexts bounds the cumulative loss incurred by the proposed sequence of denoisers from $D_k^M(x^n)$ defined in (19).

Theorem 2: For every $M > 0$, $\epsilon > 0$, $\delta > 0$, $\Delta > 0$, $1 \leq k \leq \lfloor \frac{n}{2} \rfloor$, $\exists\ \zeta^k$, ψ^k, $A^{k,M}$

$$P\left(L_{\tilde{X}^{n,M,\delta,\Delta,k}}(x^n, Y^n) - D_k^M(x^n) > \epsilon + \delta\Lambda_{\max} + \zeta^k\right) \leq A^{k,M}e^{-n\psi^k},$$

where, ζ^k, ψ^k are (as before) functions of $\mathcal{C}$, δ, Δ, ϵ, Λ and k and $A^{k,M}$ is additionally also a function of M. As with the denoiser in [2], for a given number of quantization levels, M, growth rates of $k = k_n$, $\delta = \delta_n$, $\Delta = \Delta_n$ as in [2], and $\hat{X}_{\text{univ}}^{n,M} = \tilde{X}^{n,M,\delta,\Delta,k}$ it can be shown that,

Theorem 3: For all $\mathbf{x} \in [a, b]^\infty$ and $M > 0$

$$\lim_{n\to\infty}\left[L_{\hat{X}_{\text{univ}}^{n,M}}(x^n, Y^n) - D_{k_n}^M(x^n)\right] = 0 \qquad a.s. \tag{22}$$

Additionally, for appropriate growth rates of $M = M_n$, it is also true that,

Theorem 4: For all $\mathbf{x} \in [a, b]^\infty$

$$\lim_{n\to\infty}\left[L_{\hat{X}_{\text{univ}}^{n,M_n}}(x^n, Y^n) - D_{k_n}^{M_n}(x^n)\right] = 0 \qquad a.s. \tag{23}$$

Finally, our results also imply optimality for the stochastic setting when the underlying clean signal is now a stationary process, $\mathbf{X}$, with distribution $F_{\mathbf{X}}$. Define

$$\mathbb{D}(F_{\mathbf{X}}, \mathcal{C}) = \lim_{n\to\infty}\min_{\hat{X}^n} EL_{\hat{X}^n}(X^n, Y^n) \tag{24}$$

where the expectation is assuming X^n are the first n symbols of the source with distribution $F_{\mathbf{X}}$ and the limit is guaranteed to exist by sub-additivity. Assuming $M_n \to \infty$ at a rate for which Theorem 4 holds and a quantization scheme, Q_M, that for any block length, n, and context length, k, partitions the space $\mathbb{R}^{2k}$ in a symbol-by-symbol fashion, in such a way that the resulting partition $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_{2k}$, $\{\mathcal{P}\}_{i=1}^{2k}$ being partitions in $\mathbb{R}$ corresponding to symbol-by-symbol quantization of each element in the $2k$-tuple context is asymptotically fine [8], we have

Theorem 5: For all stationary $\mathbf{X}$,

$$\lim_{n\to\infty} EL_{\hat{X}_{\text{univ}}^{n,M_n}}(X^n, Y^n) = \mathbb{D}(F_{\mathbf{X}}, \mathcal{C}) \tag{25}$$

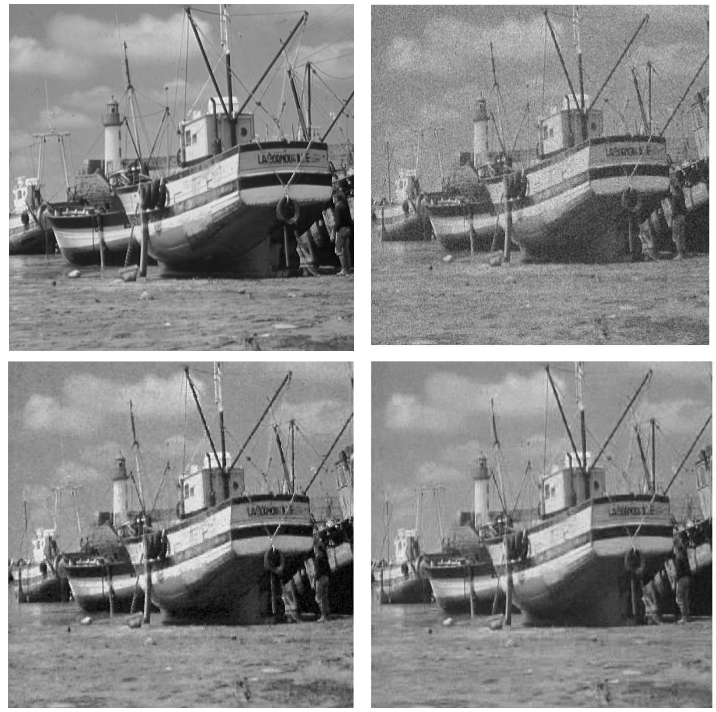

Fig. 3. Top-left: Original image, top-right: Noisy image (corrupted by an AWGN, $\sigma = 15$), bottom-left: Denoised image using the scheme in [2] ($2k+1 = 13$) RMSE = 9.441, bottom right: Denoised image using the proposed scheme RMSE = 8.913 ($M = 10$, $2k + 1 = 13$)

V. Experimental Results

Results of applying the proposed scheme to a natural test image, shown in Fig. 3. The image is corrupted by an AWGN source with $\sigma = 15$. In addition to the significant computational advantages, we achieve better denoising performance than the denoiser of [2] by considering higher order contexts, which are computationally far more tractable with this modified denoiser.

References

[1] K. Sivaramakrishnan and T. Weissman, "Universal denoising of discrete-time continuous-amplitude signals," in *"Proceedings of the IEEE International Symposium on Information Theory"*, Seattle, Jul. 2006.

[2] K. Sivaramakrishnan and T. Weissman, "Universal denoising of discrete-time continuous-amplitude signals," submitted for publication in *IEEE Transactions on Information Theory*.

[3] K. Sivaramakrishnan and T. Weissman, "Universal denoising of continuous valued signals with applications to images," in *"Proceedings of the IEEE International Conference on Image Processing"*, Atlanta, Sept. 2006.

[4] G. Motta, E. Ordentlich, I. Ramirez, G. Seroussi, and M. J. Weinberger, "The DUDE framework for continuous tone image denoising," in *"Proceedings of the IEEE International Conference on Image Processing"*, Sept. 2005.

[5] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger, "Universal discrete denoising: Known channel," *IEEE Transactions on Information Theory*, vol. 51, pp. 1229 – 1246, Jan. 2005.

[6] A. Dembo and T. Weissman, "Universal denoising for the finite input general output channel," *IEEE Transactions on Information Theory*, vol. 51, pp. 1507 – 1517, Apr. 2005.

[7] L. Devroye and L. Gyorfi, *Nonparametric Denisty Estimation, the L_1 view*, Wiley Series in Probability and Mathematical Statistics, New York, NY, 1985.

[8] L. Gyorfi and Gy. Ottucsak, "Sequential prediction of unbounded stationary time series," Preprint available at http://www.szit.bme.hu/ oti/publications/publications.htm.