1 Introduction

If we look at the DUDE algorithm from previous lectures, the loss incurred by DUDE is not much worse than that of the best $k$th order sliding window denoiser. Compared to the benchmark (the best $k$th order sliding window denoiser), is there another denoiser whose excess loss is much smaller than that incurred by the DUDE? To answer this question, we will derive a lower bound on the excess loss of any denoiser compared to the same benchmark.

2 Background and Notation

The alphabet of the noiseless signal, as well as the noisy observation and the reconstruction is a $M$-letter alphabet, denoted by $\mathcal{A}$. Denote $\prod_{i,j}$ the probability of the output symbol $j$ when the input symbol is $i$. We assume a given loss function $\Lambda : \mathcal{A}^2 \to [0, \infty)$, where $\Lambda(i,j)$ defines the loss incurred by estimating the symbol $i$ with the symbol $j$. An $n$-block denoiser is a mapping $\hat{X}_n : \mathcal{A}^n \to \mathcal{A}^n$. Let $L_{\hat{X}_n}(x^n, z^n)$ denote the normalized cumulative loss when the underlying noiseless sequence is $x^n$ and the observed sequence is $z^n \in \mathcal{A}$, i.e.,

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

A $k$-th order sliding window denoiser $\hat{X}_n$ is a denoiser that is defined by a mapping $f : \mathcal{A}^{2k+1} \to \mathcal{A}$ so that for all $z^n \in \mathcal{A}^n$

$$\hat{X}_n(z^n)[i] = f(z_{i-k}^{i+k}) , \quad i = k+1, \ldots, n-k.$$ 

Let $\mathcal{S}_k$ be the collection of all $k$th order sliding window denoiser.

**Question:** Is $k$th order DUDE in $\mathcal{S}_k$?

The answer is No. Fix a $z^n$ sequence, DUDE acts like a $k$th order sliding window denoiser for that sequence. But in general, for different $z^n$ sequences, DUDE applies different sliding window denoiser.

The $k$th order minimum loss of $(x^n, z^n)$ is defined as

$$D_k(x^n, z^n) = \min_{\hat{X}_n \in \mathcal{S}_k} L_{\hat{X}_n}(x_{k+1}^{k+n}, z^n)$$

$$= \min_{f : \mathcal{A}^{2k+1} \to \mathcal{A}} \frac{1}{n-2k} \sum_{i=k+1}^{n-k} \Lambda(x_i, f(z_{i-k}^{i+k})).$$

The expected $k$th order minimum loss is defined as

$$\hat{D}_k(x^n) \triangleq \mathbb{E}[D_k(x^n, Z^n)]$$

1
This quantity is the benchmark against which we will compare the loss incurred by other denoisers. Finally, the \( k \)th order regret \( \hat{R}_k(\hat{X}^n) \) of any \( n \)-block denoiser is defined as follows:

\[
\hat{R}_k(\hat{X}^n) = \max_{x^n \in A^n} \left( E[L_{\hat{X}_n}(a_{n-k+1}, Z^n)] - \hat{D}_k(x^n) \right)
\]  

(4)

Specifically, we know that the \( k \)th order regret of DUDE is upper-bounded by

\[
\hat{R}_k(\hat{X}^{n,k}_\text{DUDE}) \leq C \sqrt{\frac{kM^{2k}}{n}}
\]

The question we are going to answer in this lecture is whether there exists any denoiser which gives significantly better regret. We will show that for (all sufficiently large \( n \)), and any \( \hat{X}^n \)

\[
\hat{R}_k(\hat{X}^n) \geq C \frac{\alpha^k}{\sqrt{n}}.
\]

No denoiser can make regret approaching zero faster than \( \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \)

### 3 Main Result

Let \( X^n \) be a sequence of i.i.d. random variables with \( P \) denoting the distribution of \( X_i \). The quantity \( E[L_{\hat{X}_n}(X^n, Z^n)] - \hat{D}_k(X^n) \) is then a random variable. A key observation is that we can use the expectation of this random variable to lower bound the regret, i.e.,

\[
\hat{R}_k(\hat{X}^n) \geq E \left[ E[L_{\hat{X}_n}(X^n, Z^n)] - \hat{D}_k(X^n) \right] = E \left[ L_{\hat{X}_n}(X^n, Z^n) \right] - E \left[ \hat{D}_k(X^n) \right]
\]

The first term on the RHS of above equation can be lower bounded as

\[
E[L_{\hat{X}_n}(x^n, Z^n)] \geq \min_{\hat{X}^n} E[L_{\hat{X}_n}(X^n, Z^n)]
\]

The minimizer is the Bayes response, i.e.,

\[
\hat{X}^n_{\text{opt}}(z^n)[i] = \arg\min_{\hat{x}} \lambda^T_x P_{X_i|z_i}
\]

\[
= \arg\min_{\hat{x}} \lambda^T_x \frac{(P \otimes \pi_{z_i})}{P^T \pi_{z_i}}
\]

Then the optimal loss becomes

\[
D_{\text{opt}}(P) = \min_{\hat{x}} \lambda^T_x P_{X_i|z_i}
\]

Using \( D_{\text{opt}} \), the \( k \)th order regret is lower bounded by

\[
\hat{R}_k(\hat{X}^n) \geq D_{\text{opt}}(P) - E[\hat{D}_k(X^n)]
\]

(5)

#### 3.1 BSC example

\[
X_i \sim \mathbb{P} = \begin{bmatrix} 1 - P \\ P \end{bmatrix} \quad \delta < \frac{1}{2}
\]

when \( z_i = 1 \), \( \hat{X}_{\text{opt}}(z^n)(i) = \begin{cases} 0 & P < \delta \\ 1 & P > \delta \\ \text{either} & P = \delta \end{cases} \)

when \( z_i = 0 \), \( \hat{X}_{\text{opt}}(z^n)(i) = \begin{cases} 0 & P < 1 - \delta \\ 1 & P > 1 - \delta \\ \text{either} & P = 1 - \delta \end{cases} \)
We can get $\hat{X}_{opt}$ and $D_{opt}(P)$ by combining two cases

$$\hat{X}_{opt} = \begin{cases} 
"always say 0" & P \leq \delta \\
"say what you see" & \delta \leq P \leq 1 - \delta \\
"always say 1" & P \geq 1 - \delta 
\end{cases}$$

$$D_{opt}(P) = \begin{cases} 
P & P \leq \delta \\
\delta & \delta \leq P \leq 1 - \delta \\
1 - P & P \geq 1 - \delta
\end{cases}$$

Observe that when $P = \delta$, the crossover probability, there are two Bayes optimal denoisers, namely, the "always say o" and the "say-what-you-see" denoiser. We will try to lower bound the regret for $P = \delta$.

when $P = \delta$, \begin{equation}
\hat{R}_k(\hat{X}^n) \geq \delta - E[\min_{\hat{X} \in S_0} L_{\hat{X}}(X^n, Z^n)]
\end{equation}

The second term of the RHS of the above equation can be handled as follows:

$$E \left[ \min_{\hat{X} \in S_0} L_{\hat{X}}(X^n, Z^n) \right] \leq E \left[ \min \{ L_{\text{always 0}}(X^n, Z^n), L_{\text{swys}}(X^n, Z^n) \} \right]$$

$$= E \left[ \min \left\{ \frac{1}{n} \sum_{i=1}^{n} 1\{X_i = 1\}, \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \neq Z_i\} \right\} \right]$$

$$= \delta + \frac{1}{\sqrt{n}} E \left[ \min \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1\{X_i = 1\} - \delta), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1\{X_i \neq Z_i\} - \delta) \right\} \right]$$

Note that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1\{X_i = 1\} - \delta)$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1\{X_i \neq Z_i\} - \delta)$ are sums of independent random variables. Further each set of random variables are independent of each other. Therefore, by the Central Limit Theorem, they coverge in distribution to independent zero mean Gaussian random variables ($N(0, \delta(1 - \delta))$).

Therefore,

$$E[\min_{\hat{X} \in S_0} L_{\hat{X}}(X^n, Z^n)] \approx \delta - \frac{C}{\sqrt{n}}, \quad C > 0.$$ 

Now we have

$$\hat{R}_k(\hat{X}^n) \geq \delta - E[\min_{\hat{X} \in S_0} L_{\hat{X}}(X^n, Z^n)]$$

$$\geq \delta - (\delta - \frac{C}{\sqrt{n}})$$

$$= \frac{C}{\sqrt{n}}$$

(6)

### 3.2 Proof of Lower bound

Definition: $(\pi, \lambda)$ is *neutralizable* if $\exists$ channel output symbol $t \in A$, s.t. for some $P \in M$ ($M$ : simplex $M$-dimensional)

1. $\lambda_t^T (P \circ \pi_t) = \lambda_t^T (P \circ \pi_t) = \min_{\pi} \lambda_t^T (P \circ \pi_t)$
2. $(\lambda_i - \lambda_j) \circ P \circ \pi_t \neq 0$

Further the distribution $P$ that satisfies the two equations is termed *loss-neutral*.
Theorem 1. : For any neutralizable \((\pi, \lambda)\), and any sequence of denoisers \(\hat{X}^n\)

\[
\hat{R}_k(\hat{X}^n) \geq \frac{C}{\sqrt{n}} \left( \sum_a \sqrt{(P^*)^T \pi_a} \right)^2 (1 + o(1))
\]

where \(P^*\) is any loss-neutral distribution and \(C\) is a positive function of \((\pi, \lambda)\) and \(P^*\).

Proof: Let

\[
q(z^n, x^n, c^k_{-k})[\alpha] = \frac{\{|i : z_i^{k} = c^k_{-k}, x_i = \alpha\}|}{n - 2k}, \text{ where } \alpha \in \mathcal{A}.
\]

Suppose \(X^n\) is iid with \(X_i \sim \mathbb{P}\)

\[
E[q(z^n, x^n, c^k_{-k})] = \mathbb{P}(\mathbb{P} \circ \pi_0) \prod_{i=-k}^{k} \mathbb{P}^x \pi_{c_i}
\]

\[
\begin{align*}
D_k(x^n, z^n) &= \min_{f: \mathcal{A}^{k+1} \rightarrow \mathcal{A}} \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} \Lambda(x_i, f(z_i^{k})) \\
&= \sum_{c^k_{-k} \in \mathcal{A}^{k+1}} \min_{\tilde{x} \in \mathcal{A}} \sum_{j} \Lambda(j, \tilde{x}) q(z^n, x^n, c^k_{-k})[j] \\
&= \sum_{c^k_{-k} \in \mathcal{A}^{k+1}} \min_{\tilde{x} \in \mathcal{A}} \sum_{\beta \in \mathcal{A}} \Lambda(\tilde{x}, \beta) q(z^n, x^n, c^k_{-k})
\end{align*}
\]

**Definition:** \(X_1, \ldots, X_n\) is \(m\)-dependent if for all \(s > r + m\), \(X_1, \ldots, X_r\) and \(X_{s}, \ldots, X_n\) are independent

**Theorem 2.** For a stationary \(m\)-dependent sequence \(X^n\) s.t. \(E[X_i] = 0\) and \(E[|X_i|^3] < \infty\),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \rightarrow N(0, V), \text{ as } n \rightarrow \infty,
\]

where \(V = E[X_i^2] + 2 \sum_{k=2}^{m+1} E[X_1, X_k]\)

This theorem is proved by Hoeffding and Robbins [1]. The following lemma is a consequence of the theorem. The proof of the lemma can be found in [2].

**Lemma 3.** \(X^n\) is iid and \(X_i \sim \mathbb{P}\). Then, for any \(\alpha \in \mathbb{R}^n\) and any \(c^k_{-k} \in \mathcal{A}^{k+1}\) s.t. \(\alpha^T (\mathbb{P} \circ \pi_0) = 0\),

\[
\lim_{n \rightarrow \infty} \mathbb{P} \left[ \sqrt{n} |\alpha^T q(z^n, x^n, c^k_{-k})| \right] = \sqrt{\frac{\alpha V}{\pi}}
\]

where \(V = (\alpha \circ \alpha)^T (\mathbb{P} \circ \pi_0) \prod_{i=-k}^{k} \mathbb{P}^x \pi_{c_i} - \beta \mathbb{P}^x \pi_{c_i} \).

Let us complete the proof of Theorem 1.

Recall Eq. 5,

\[
\hat{R}_k(\hat{X}^n) \geq D_{\text{opt}}(\mathbb{P}) - E[\hat{D}_k(X^n)]
\]

If \(X_i \sim \mathbb{P}^*\) loss-neutral w.r.t. \((\pi_i, \lambda_i, \lambda_j)\),

\[
\hat{R}_k(\hat{X}^n) \geq \sum_{c_0 \in \mathcal{A}} \min_{\tilde{x} \in \mathcal{A}} \Lambda(\tilde{x}) (\mathbb{P}^* \circ \pi_0) - E[\hat{D}_k(X^n)]
\]

(7)
Let us find an upper bound for the second term of R.H.S. of the above equation.

\[
\mathbb{E}[\hat{D}_k(X^n)] = \mathbb{E}[D_k(X^n, Z^n)] = \mathbb{E} \left[ \sum_{c_{-k}^k} \min_{x \in A} \lambda_T^T q(Z^n, X^n, c_{-k}^k) \right]
\]

\[
= \sum_{c_{-k}^k, c_0 \neq t} \mathbb{E}[\min_{x \in A} \lambda_T^T q(Z^n, X^n, c_{-k}^k)] + \sum_{c_{-k}^k, c_0 = t} \mathbb{E}[\min_{x \in A} \lambda_T^T q(Z^n, X^n, c_{-k}^k)]
\]

The first term on the R.H.S. of Eq. 8 can be upper bounded as follows:

\[
\sum_{c_{-k}^k, c_0 \neq t} \mathbb{E}[\min_{x \in A} \lambda_T^T q(Z^n, X^n, c_{-k}^k)] \leq \sum_{c_{-k}^k, c_0 \neq t} \min_{x \in A} \mathbb{E}[\lambda_T^T q(Z^n, X^n, c_{-k}^k)]
\]

\[
= \sum_{c_{-k}^k, c_0 \neq t} \min_{x \in A} \lambda_T^T (P \circ \pi_{c_0}) \prod_{i = -k, i \neq 0}^{k} P_T \pi_{c_i}
\]

The second term on the R.H.S. of Eq. 8 can be upper bounded in the following way:

\[
\sum_{c_{-k}^k, c_0 = t} \mathbb{E} \left[ \min_{x \in A} \lambda_T^T q(Z^n, X^n, c_{-k}^k) \right] \\
\leq \sum_{c_{-k}^k, c_0 = t} \mathbb{E} \left[ \min \{\lambda_i^T q(Z^n, X^n, c_{-k}^k), \lambda_j^T q(Z^n, X^n, c_{-k}^k)\} \right]
\]

\[
= (a) \frac{1}{2} \sum_{c_{-k}^k, c_0 = t} \left( \mathbb{E}[\lambda_i^T q(Z^n, X^n, c_{-k}^k)] + \mathbb{E}[\lambda_j^T q(Z^n, X^n, c_{-k}^k)] - (\lambda_i - \lambda_j)^T q(Z^n, X^n, c_{-k}^k) \right)
\]

\[
= (b) \min_{x \in A} \lambda_T^T (P \circ \pi_t) - \frac{1}{2} \sum_{c_{-k}^k, c_0 = t} \mathbb{E}[((\lambda_i - \lambda_j)^T q(Z^n, X^n, c_{-k}^k)]
\]

\[
= (c) \min_{x \in A} \lambda_T^T (P \circ \pi_t) - \sum_{c_{-k}^k, c_0 = t} \frac{C(1 + o(1))}{\sqrt{n}} \sqrt{\frac{2V_{c_{-k}}}{\pi}}
\]

where \(V_{c_{-k}} = ((\lambda_i - \lambda_j) \circ (\lambda_i - \lambda_j)) T (P \circ \pi_{c_0}) \prod_{i = -k, i \neq 0}^{k} P_T \pi_{c_i}\), (a) follows because \(\min\{x, y\} = \frac{x + y - |x - y|}{2}\), (b) follows from the definition of neutralizable and loss-neutrality, and (c) is due to Lemma 3.

Finally, substituting Eq. 9 and Eq. 10 into Eq. 8 and then using the obtained upper bound of \(\mathbb{E}[\hat{D}_k(X^n)]\) in Eq. 7, we have

\[
\hat{R}_k(X^n) \geq D_{opt}(P) - \mathbb{E}[\hat{D}_k(X^n)]
\]

\[
= \sum_{c_{-k}^k} \min_{x} \lambda_T^T (P \circ \pi_t) - \sum_{c_{-k}^k} \min_{x} \lambda_T^T (P \circ \pi_t) + \frac{C(1 + o(1))}{\sqrt{n}} \sum_{c_{-k}^k, c_0 = t} \sqrt{\frac{2V_{c_{-k}}}{\pi}}
\]

\[
= \sum_{c_{-k}^k, c_0 = t} \frac{C(1 + o(1))}{\sqrt{n}} \sqrt{\frac{2V_{c_{-k}}}{\pi}}.
\]
Observe that
\[
\sum_{c^k \in A^{2k+1}, c_0=1} \frac{2V_{c^k}}{\pi} = \sqrt{\frac{2}{\pi}} \sqrt{\left(\sum_{a^k \in A^{2k}} \left(\prod_{i=1}^{2k} (P^* \pi_{a_i}) \right) \right)^2}.
\]

Note that since \(P^*\) is a loss-neutral distribution \((\lambda_i - \lambda_j) \odot P^* \odot \pi_t \neq 0\), and therefore
\[
((\lambda_i - \lambda_j) \odot (\lambda_i - \lambda_j))^T (P \odot \pi_t) > 0.
\]

Also observe that
\[
\sum_{a^k \in A^{2k}} \left(\prod_{i=1}^{2k} (P^* \pi_{a_i}) \right)^{1/2} = \left(\sum_{a \in A} \sqrt{(P^* \pi_a)^T} \right)^{2k}.
\]

For more details and discussions see [2].

References
