

Universal Discrete Denoising: Known Channel

Tsachy Weissman^{1,3}, Erik Ordentlich¹, Gadiel Seroussi¹, Sergio Verdú² and Marcelo Weinberger¹

emails: tsachy@stanford.edu, eord@hpl.hp.com, seroussi@hpl.hp.com, verdu@princeton.edu, marcelo@hpl.hp.com

Finite-alphabet signals corrupted by discrete noisy channels arise naturally in a wide range of applications spanning many fields of science and engineering. In the Shannon paradigm, redundancy is added to the noiseless signal in order to protect it from the channel noise, and a decoder that knows the codebook and the channel statistics can recover the noiseless signal with arbitrary reliability, provided that the coding scheme respects the fundamental limits of information theory. In contrast, there are many applications where no channel coding is performed and the recovery of the corrupted signal can only be accomplished with a certain distortion. Examples include typing correction, hidden Markov model state estimation, DNA sequence analysis, image enhancement, pattern recognition, blind equalization, and joint source-channel decoding, to name a few. A *universal* discrete denoiser is a scheme that (asymptotically) attains the optimum distribution-dependent performance with no a priori knowledge of the source statistics. Despite the importance of the problem, there is no universal discrete denoiser available in the literature. Furthermore, some of the denoising schemes that have been suggested for the universal setting are not practical in that their complexity grows exponentially with the size of the noisy data. The main contribution of this work is a discrete universal denoiser (DUDE) that attains optimum (distribution-dependent) performance with computational complexity that grows linearly with the data size.

We shall assume (cf. [1] for a discussion of the more general case) that the components of the clean, as well as of the noise-corrupted signal, take their values in the same M -ary alphabet $\mathcal{A} = \{1, \dots, M\}$, that the corruption mechanism is a discrete memoryless channel (DMC) represented by the matrix $\Pi = \{\Pi(i, j)\}_{i, j \in \mathcal{A}}$ ($\Pi(i, j)$ denoting the probability of a channel output symbol j when the input is i) which is known and invertible, and that there is a given loss function $\Lambda : \mathcal{A}^2 \rightarrow [0, \infty)$, represented by the matrix $\Lambda = \{\Lambda(i, j)\}_{i, j \in \mathcal{A}}$ ($\Lambda(i, j)$ denoting the loss incurred by estimating the symbol i with the symbol j). We let π_i denote the i -th column of Π , and λ_j denote the j -th column of Λ . Hence, $\Pi = [\pi_1 | \dots | \pi_M]$, $\Lambda = [\lambda_1 | \dots | \lambda_M]$. An n -block denoiser is a mapping $\hat{X}^n : \mathcal{A}^n \rightarrow \mathcal{A}^n$. We now describe the DUDE: For $2k < n$ and $\mathbf{u}_l, \mathbf{u}_r \in \mathcal{A}^k$ let $\mathbf{m}(z^n, \mathbf{u}_l, \mathbf{u}_r)$ denote the M -dimensional column vector whose β -th component, $\beta \in \mathcal{A}$, is given by

$$\left\{ i : k+1 \leq i \leq n-k, z_{i-k}^{i-1} = \mathbf{u}_l, z_i = \beta, z_{i+1}^{i+k} = \mathbf{u}_r \right\}.$$

Let $\hat{X}^{n,k}$ denote the n -block denoiser defined by

$$\hat{X}^{n,k}(z^n)[i] = \arg \min_{\hat{x} \in \mathcal{A}} \mathbf{m}^T(z^n, z_{i-k}^{i-1}, z_{i+1}^{i+k}) \Pi^{-1} [\lambda_{\hat{x}} \odot \pi_{z_i}],$$

for $k+1 \leq i \leq n-k$ (the value at $i \leq k$ and $i > n-k$ is arbitrarily assigned), where $\hat{X}^{n,k}(z^n)[i]$ denotes the i -th coordinate of $\hat{X}^{n,k}(z^n)$. Finally, for a given sequence $\{k_n\}$ (which will be specified below) we let, for each n , $\hat{X}_{\text{univ}}^n = \hat{X}^{n,k_n}$. A natural implementation of the denoising algorithm makes two passes through the observations z^n . The counts $\mathbf{m}(z^n, \mathbf{u}_l, \mathbf{u}_r)$ are accumulated and stored in the first pass while the actual denoising is performed in the second pass. This can be shown to require a linear number of register-level operations and sub-linear working storage size relative to the input data length.

In order to state the optimality results of the DUDE we need the following definitions: Let $L_{\hat{X}^n}(x^n, z^n)$ denote the normalized cumulative loss, as measured by Λ , of the denoiser \hat{X}^n when the observed sequence is $z^n \in \mathcal{A}^n$ and the underlying one is $x^n \in \mathcal{A}^n$, i.e., $L_{\hat{X}^n}(x^n, z^n) = \frac{1}{n} \sum_{i=1}^n \Lambda(x_i, \hat{X}^n(z^n)[i])$. We define the k -th order sliding-window minimum loss by

$$D_k(x^n, z^n) = \min_{f: \mathcal{A}^{2k+1} \rightarrow \mathcal{A}} \left[\frac{1}{n-2k} \sum_{i=k+1}^{n-k} \Lambda(x_i, f(z_{i-k}^{i+k})) \right].$$

By *Semi-Stochastic Setting* we refer to the case where \mathbf{x} is an individual sequence and \mathbf{Z} is its noise-corrupted version, i.e., \mathbf{Z} is the output of the memoryless channel, Π , whose input is \mathbf{x} . In the statement of the following result we assume the semi-stochastic setting.

Theorem 1 For all $\mathbf{x} \in \mathcal{A}^\infty$, the sequence of denoisers $\{\hat{X}_{\text{univ}}^{k_n}\}$ with $\lim_{n \rightarrow \infty} k_n = \infty$ satisfies

$$\lim_{n \rightarrow \infty} [L_{\hat{X}_{\text{univ}}^{k_n}}(x^n, Z^n) - D_{k_n}(x^n, Z^n)] = 0 \quad \text{a.s.},$$

provided $k_n M^{2k_n} = o(n/\log n)$.

Assume now that the noiseless signal is a stationary stochastic process \mathbf{X} . Letting $P_{X^n}, P_{\mathbf{X}}$ denote, respectively, the distributions of X^n, \mathbf{X} , and \mathcal{D}_n denote the class of all n -block denoisers, we define $\mathbf{D}(P_{X^n}, \Pi) = \min_{\hat{X}^n \in \mathcal{D}_n} EL_{\hat{X}^n}(X^n, Z^n)$, the expectation on the right side assuming that $X^n \sim P_{X^n}$ and that Z^n is the noisy observation of X^n (corrupted by the DMC Π). Define now $\mathbf{D}(P_{\mathbf{X}}, \Pi) = \lim_{n \rightarrow \infty} \mathbf{D}(P_{X^n}, \Pi)$ (where the limit can be shown to exist by subadditivity). By definition, $\mathbf{D}(P_{\mathbf{X}}, \Pi)$ is the (distribution-dependent) optimum asymptotic denoising performance attainable when the noiseless signal is emitted by the source $P_{\mathbf{X}}$ and corrupted by the channel Π .

Theorem 2 The sequence of denoisers $\{\hat{X}_{\text{univ}}^{k_n}\}$ with $\lim_{n \rightarrow \infty} k_n = \infty$ satisfies, for all stationary processes \mathbf{X} ,

$$\lim_{n \rightarrow \infty} EL_{\hat{X}_{\text{univ}}^{k_n}}(X^n, Z^n) = \mathbf{D}(P_{\mathbf{X}}, \Pi),$$

provided $\sqrt{k_n} M^{k_n} = o(\sqrt{n})$.

REFERENCES

- [1] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger, "Universal discrete denoising: Known channel," *HP Labs Tech. Report*, HPL-2003-29, February 2003. Also submitted to *IEEE Trans. Inform. Th.* (manuscript available at: <http://www.hpl.hp.com/infotheory/dude/index.htm>).

¹Hewlett-Packard Laboratories, Palo Alto, CA 94304, USA
²Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
³Department of Statistics, Stanford University, Stanford, CA 94305, USA